

Ján Šefránek: Inteligencia ako výpočet

IRIS, Bratislava 2000, 427 s.

Umělá inteligence (AI) je velice široká problematika. Je to oblast, ve které se setkáváme s nejrozmanitějšími přístupy k lidským aktivitám, a pokud pojímáme umělou inteligenci jako vědu, pak její charakter je nesporně interdisciplinární a mnohotvárný. Jak říká autor v předmluvě:

Různé oblasti umelej inteligencie (a přístupy k nej) sa výrazne líšia v závislosti od používanej konceptuálnej a metodologickej výbavy, od zamerania pozornosti, priorit, optiky a cieľov. Oblasti umelej inteligencie sú tak pestro diferencované, že si často nemajú vzájomne čo povedať. Presne ako ľudia. Preto nie je možné (domnievam sa) hovoriť o základoch umelej inteligencie ako celku.

Přesto bývá často charakterizován hlavní cíl výzkumu v oblasti umělé inteligence jako „naučit stroje myslet“. Otázkou, zdali se vůbec jedná o problém řešitelný, zda tedy může stroj myslet, se zabývalo mnoho badatelů již od nepaměti, ovšem její relevance se stává obzvláště důležitou zejména s rozvojem počítačů v posledních padesáti letech. Pomíneme-li příliš zjednodušený (sci-fi) pohled na inteligenci, který vedl k naivním optimistickým představám šedesátých let o inteligentních robotech-počítačích, které prožijí své dětství, naučí se jazyk, získají dokonalé znalosti o světě a zcela ovládnou doménu lidského myšlení, můžeme snad přesněji charakterizovat výzkum v oblasti AI jako snahu vytvořit *model* lidského myšlení. Dnes je snad již zřejmé, že tento model nebude nikdy např. *moudrým* psychiatrem či soudcem. Proto asi nejlepší definice umělé inteligence pochází od M. Minského (MIT):

Umělá inteligence je věda o vytváření strojů využívajících postupy, které prováděny člověkem by vyžadovaly jeho inteligenci.

Tyto člověkem vytvořené stroje by pak měly mít schopnost emulovat a simulovat lidské metody pro deduktivní a induktivní akvizici a aplikaci znalostí a odvozování.

Z hlediska použitých principů fungování může být umělá inteligence rozdělena na symbolicky orientované systémy a „analogově“ orientované systémy (neuronové sítě). Kniha J. Šefránka se nezabývá neuronovými sítěmi. Text je věnován *teoretickým základům znalostního paradigmatu* umělé inteligence, tedy přístupu, jehož cílem je výpočtově (komputačně) modelovat zpracování poznatků, které jsou reprezentovány symbolicky větami nějakého jazyka, a pochopit (především hypotetické) usuzování. Z hlediska umělé inteligence jde o studium tzv. *nemonotónního usuzování*, tj. usuzování, které je doprovázeno revizí závěrů či hypotéz, přičemž tyto hypotézy byly odvozeny na základě neúplných nebo i nekonzistentních poznatků, nebo obecněji jde o studium hypotetického, na kontextu závislého usuzování a jeho dynamických aspektů.

Pozornost je zaměřena na „počítání“ s větami, poznatky, tedy na výpočtové modelování způsobů lidské argumentace. Jelikož je umění správné argumentace snad nejvlastnějším předmětem logiky, je tato teorie ovlivněna především výzkumem v oblasti logiky a speciálně logického programování, avšak také např. výzkumem deduktivních databází, problémy vyčísitelnosti a složitosti výpočtů, apod. Umělá inteligence, jakožto interdisciplinární věda, využívá a svým způsobem integruje výsledky práce (např.) filosofů, logiků, matematiků, lingvistů, ale i psychologů či neurofyziologů. Snad nejtěsnější vazby a návaznosti je však možno nalézt v oblasti logiky (filozofické i matematické), které se v podstatě primárně zabývají tvorbou formálních modelů usuzování a poznání. Umělá inteligence se od nich liší snad pouze důrazem na výpočtové aspekty.

Záběr knihy je, tak jako celá oblast AI, velice široký (o čemž svědčí i nemalý rozsah textu – 427 stran). Autorovi jde však především o teoretický výklad uvedené problematiky zaměřený na nejmodernější, u nás dosud málo známé poznatky. Nenajdeme zde proto informace o komerčních produktech ani návody na tvorbu takových aplikací. Autor tedy (zřejmě záměrně) zcela opomíjí např. problematiku expertních systémů a plánovacích systémů, a také se kromě logických programovacích jazyků nevěnuje jiným programovacím jazykům používaným v AI, jako jsou funkcionální jazyky (zejména LISP) a objektové orientované jazyky.

V první kapitole **Východiská** jsou formulovány základní principy, motivace a intuice, na kterých je postaven text celé práce. Tyto odstavce jsou sice formulovány poněkud vágně, avšak jsou důležité a relevantní pro pochopení problematiky, neboť vymezují předmět zkoumání a zasazují jej do širších souvislostí. Za výpočtový model inteligentního chování je tedy považován program (nebo spíše jeho formální specifikace), který jistým způsobem vyhovuje charakteristikám tohoto chování. Naše pozornost bude přitom zaměřena pouze na *znalostní systémy*, tj. programové systémy, které sestávají ze dvou komponent: *báze znalostí a inferenčního stroje*. Báze znalostí je reprezentována množinou E vět nějakého formálního jazyka, inferenční stroj je realizován operátorem odvození Cn, přičemž odvození nemusí být (a často nebude) pouze deduktivní. Pozornost však není soustředěna na syntaktické problémy, nýbrž je z velké části orientována na sémantiku. Zde je důležitá autorova poznámka 9 na str. 19, kde upřesňuje, jak bude v tomto textu chápána sémantika: „Sémantiku nějakého jazyka definujeme tehdy, když jazykovým výrazům přiřazujeme nějaké objekty, kterými mohou být nějaké formální konstrukce, dokonce i jazykové výrazy. Důležité je právě toto přiřazení, ne povaha přiřazených objektů.“ Jistě, takovýto formalistický, či interpretační přístup k sémantice je dnes běžný. Nutno však poznamenat, že není jediný, a zejména při logické analýze přirozeného jazyka neprávě nevhodnější, jak ukazuje např. Pavel Tichý v mnoha svých článcích a ve své knize „The Foundations of Frege's Logic“.

Kapitola 2 **Databáze** pojednává v první části o klasických relačních databázích, neboť relační databáze můžeme vlastně považovat za jednodušší typ znalostních systémů, kde znalosti jsou reprezentovány ve formě tabulek (ze kterých generujeme příslušné věty báze znalostí E) a operátor odvození Cn je realizován jednotlivými databázovými dotazovacími jazyky, s více či méně dobře definovanou sémantikou,

např. relačním kalkulem či relační algebrou. Přitom relační algebra je považována za sémantiku relačního kalkulu. Ve druhé části této kapitoly jsou naznačeny možnosti zesílení a zevšeobecnění databázového jazyka pro databáze s neúplnými nebo nejistými hodnotami. Jde v podstatě o zavedení nulových hodnot, které reprezentují stav databáze, kdy hodnota určitého atributu neexistuje nebo je neznámá. Zde se setkává čtenář s důležitou sémantickou strukturou. Jde o speciální případ Kripkeho struktury, která je zde definována jako trojice (W, r, ρ) , kde W (množina možných světů) je množina možných interpretací (relací bez nulových hodnot) jistého databázového predikátu p , r je relace s nulovými hodnotami, která rovněž interpretuje predikát p , a ρ je relace dostupnosti mezi prvky W definovaná tak, že z relace r jsou dostupné všechny relace, které jsou úplněním relace r (tj. nulové hodnoty jsou nahrazeny jistými možnými hodnotami příslušného typu). Autor dále naznačuje možnosti, jak vyhodnocovat selekci relační algebry. Tyto možnosti v podstatě odpovídají dvěma extrémům (*nutně* – ve všech „možných světech“ dostupných z relace r , a *možná* – v některých „možných světech“ dostupných z relace r). Kripkeho struktury jsou v knize pojednány znovu v kap. 7.3, avšak zde již autor v podstatě předpokládá obeznámenost s tímto pojmem. Rovněž zmínka o možných světech v takto zúženém pohledu je zřejmě spíše zavádějící než informativní. Proto by možná bylo žádoucí, kdyby autor o těchto strukturách pojednal obecněji a šířeji, např. v některém dodatku, a to zejména v souvislosti se sémantikou *modálních logik*. Třetí část této kapitoly se zabývá deduktivními databázemi, kde je specifikace databázového jazyka založena na jazyku logického programování. Autor zde charakterizuje datalogovské databáze.

Ve třetí kapitole – **Databáze s hypotézami** – se čtenář v podstatě poprvé v textu setkává s nemonotónním usuzováním. Databáze s hypotézami totiž obsahuje tvrzení (pravidla), která jsou většinou pravdivá, „až na nějaké výjimky“. Inferenční stroj je zde generátor hypotéz a odvození již nemusí být deduktivní a monotónní. Závěry, tj. hypotézy přijímáme, dokud nedojde při asimilaci nových poznatků k nekonzistenci. Některé závěry jsou tedy zařazeny do báze poznatků „natrvalo“, jiné je možno po čase odmítnout, když přijmeme poznatky, kterým důvěřujeme více a které jsou s dřívějšími hypotézami v rozporu. Problému revize odvozených závěrů či dříve získaných poznatků je vlastně věnován celý zbytek knihy. Klasickým příkladem nemonotónního (nekorektního) operátoru odvození je tzv. „předpoklad uzavřeného světa“, neboli (v logickém programování) „negace jako konečné selhání“. Důležitá je zde *základní charakteristika* hypotetického usuzování: Operátor C_n , který je reflexivní (tj. $X \subseteq C_n(X)$) pro každou množinu vět X) a není logicky korektní, přičemž zachovává konzistenci, je *nemonotónní*. Klasickým příkladem je odvozování generalizací. Mějme bázi poznatků E , která obsahuje tisíc vět: „labuť₁ je bílá“, „labuť₂ je bílá“, ..., „labuť₁₀₀₀ je bílá“, tedy $C_n(E)$ obsahuje E plus zobecnění „všechny labutě jsou bílé“. Tisící první poznatek ϕ však je „tato labuť je černá“. Kdyby byl operátor C_n monotónní, tedy kdyby bylo $C_n(E) \subseteq C_n(E \cup \phi)$, byla by naše množina $C_n(E \cup \phi)$ hypotéz nekonzistentní. Tedy odmítneme monotónnost a provedeme revizi: Implicitní poznatek „všechny labutě jsou bílé“ již nebude prvkem $C_n(E \cup \phi)$. Přitom nás vlastně

nic, kromě „zdravého selského rozumu“, nenutí k nemonotónnosti. Problém, jak ošetřit nekonzistenci, může být řešen také jiným způsobem než revizí (což je většinou velmi „drahá operace“). Nekonzistence nám v podstatě vadí proto, že operátor C_n je *triviální*, tj. zachovává princip, že „ze sporu je možno odvodit cokoli“. Tedy je-li A nekonzistentní množina, pak $C_n(A)$ je množina *všech* vět daného jazyka. Proto druhým možným přístupem, jak ošetřit nekonzistence, je vzdát se triviálnosti operátoru C_n (neboť operátor C_n , který je triviální a nezachovává konzistenci, se již chová naprosto defektně: i z konzistentní množiny odvodí vše) a pokračovat v „rozumném“ netriviálním odvozování. Z praktického hlediska jde o opatrnou inferenci i za přítomnosti nekonzistencí kombinovanou s tím, že revize odložíme na později (např. až budeme mít více informací-podkladů pro rozhodnutí, co vlastně revidovat, „škrtnout“). Autor tedy v odstavci 3.2. velice pěkným způsobem předkládá motivace a intuice, které stojí v pozadí využití tzv. *parakonzistentních logik* ve znalostních systémech a stručně charakterizuje netradiční principy parakonzistentních logik (čtyř hodnotová interpretace a kvazi-vyplyvání), přičemž také upozorňuje na důsledky, které jsou nepříjemné z logického hlediska (např. neplatí *modus ponens*). V závěrečném odstavci této kapitoly stojí za zmínku autorovo poněkud neopatrné tvrzení, že

V počiatkoch umelej inteligencie sa veľká pozornosť venovala klasickej logike a automatickému dokazovaniu teorém. Časom sa však zistilo, že to nie sú adekvátne nástroje na modelovanie inteligencie.

Takováto tvrzení vedou často k ukvapeným závěrům, že klasická logika se svými pojmy logického vyplývání, teorému apod. není v praxi využitelná, což autor jistě nechtěl říci. Vždyť bez těchto klasických pojmů bychom vůbec nebyli schopni formulovat žádnou rozumnou teorii umělé inteligence a jsou tedy základním stavebním kamenem umělé inteligence. Že nejsou jediným a že jejich „implementace“ není vždy neefektivnější, je snad samozřejmé. Ostatně problém efektivnosti hypotetického usuzování zde bude před námi stát zřejmě vždy, neboť usuzování je náročná činnost a její modelování bude vždy výpočtově náročné.

Jelikož jsou první tři kapitoly knihy svým obsahem fundamentální pro pochopení mnoha aspektů umělé inteligence a celého obsahu knihy, věnovala jsem jim značnou pozornost. Další kapitoly, tj. **4 Typické případy a hierarchie**, **5 Generalizácie a vysvetlenia**, **6 Sémantika falzifikácie**, **7 Dynamika poznania a usudzovania a 8 Protiv bariere nezávládnutelnosti** již nebudu tolik komentovat, jen stručně zrekapituluji jejich obsah. Tím ovšem nechci naznačit, že by byly jakkoli méně důležité. Naopak, je to jádro knihy, její cenný přínos, avšak rozsah této recenze je limitovaný. Kapitola 4 nás uvádí do problematiky „Defaultových teorií“ (s pravidly, která neplatí vždy, a proto obsahují strážce proti zavlečení nekonzistence) a „Hierarchických sítí“. Zde je velice zajímavá ukázka toho, jak je možno zefektivnit klasické rezoluční dokazování zavedením hierarchicky uspořádaných typů. Autor rovněž vyvrací jeden z mýtů umělé inteligence, že v jazyce logiky není možno reprezentovat to, co v hierarchických sítích. K této části snad ještě jedna poznámka. Vztah ISA mezi třídami je klasicky v teorii databází definován jako *nutný* vztah, v AI tomu však tak není, zde ISA znamená *obvykle*, což pak má samozřejmě jisté nepříjemné důsledky,

např. dědění konfliktních vlastností z více zdrojů. Autor se bohužel o tomto rozdílu v chápání ISA nezmiňuje. Kapitola 5 se věnuje indukci a abdukci, tedy způsobům generalizace a vysvětlení. Indukce (generalizace) je jistým způsobem inverzní ke klasické dedukci, abdukce je pak výpočtovým modelováním vysvětlování, což může pomoci řešit problémy jako stanovení diagnózy poruchy, problém plánování, avšak rovněž např. problém porozumění přirozenému jazyku (jaký je možný význam věty při jejím užití v daném kontextu). Právě tento poslední velice zajímavý problém je však pouze zmíněn a autor se mu nijak dále nevěnuje. Kapitola 6 je v podstatě výkladem teorie logického programování. Odstavec 6.1 pojednává o metodách stratifikace (rozvrstvení) – stratifikovaný program má standardní model a rozumnou sémantiku. V odstavci 6.2 jsou definovány stabilní a dobře fundované modely logického programu, odstavec 6.3 řeší problém složitosti a implementace, v odstavci 6.4 autor ukazuje, jak je možno využít sémantiky negace při reprezentaci znalosti a konečně odstavec 6.5 stručně shrnuje výsledky Przymusińského, které se dají formulovat asi takto: Pro stratifikované logické programy je sémantika perfektních modelů ekvivalentní se všemi čtyřmi formalizmy nemonotónního usuzování, tj. s defaultovou teorií, předpokladem uzavřeného světa, cirkumskripcí a autoepistemic-kou logikou. Kapitola 7 se nejprve v odstavci 7.1, podrobněji zabývá problémem revize, tedy dynamiky poznání a usuzování. Dále je v odstavci 7.2. pojednáno o problému sjednocujících pohledů na různé přístupy, které byly až dosud popsány. Nakonec je v odst. 7.3. představen podrobněji jeden takovýto sjednocující pohled, který je založen na dynamických Kripkeho strukturách (DKS). Zde autor popisuje velice zajímavou teorii AELKB (autoepistemická logika poznání a víry, tj. teorie postojů ideálně introspektivního agenta) a využívá DKS k sémantické charakterizaci této teorie. Jak jsem již zmínila na začátku této recenze, Kripkeho struktury a zejména DKS by si „zasloužily“ další dodatek, také proto, že se zde autor dopouští jistých nepřesných formulací. Tak např. na str. 286 definuje „úplný S5-rámec“ jako Kripkeho strukturu (W, ρ) , kde W je množina možných světů a $\rho = W \times W$. Vzápětí však říká, že možnými světy budou úplné S5-rámce. Dále na str. 287 definuje zobrazení m , které každému možnému světu w přiřadí nějakou interpretaci m_w . Avšak na začátku odstavce charakterizoval množinu možných světů právě jako množinu mnohých (všech) možných (naprosto „formálních“) interpretací jazyka (což je mimochodem neobvyklé pojetí, za zmínku by jistě stálo alespoň klasické Leibnizovo pojetí nebo moderní Tichého pojetí). Na str. 285-6 říká, že je rozumné očekávat, že agent pokládá $K\phi$ ($B\phi$) za pravdivé tehdy, když *ví*, že ϕ (nebo *věří*, že ϕ) nezávisle na tom, zda ϕ je pravdivé nebo nepravdivé. Tim ovšem zcela opomíjí zásadní rozdíl mezi vědět a věřit, který spočívá v tom, že vědět, na rozdíl od věřit, je *faktivum* (tedy z $K\phi$ vyplývá pravdivost ϕ). Poznámka 14 pod čarou celou věc spíše ještě více zatemní. V závěrečné kapitole 8 pak autor nastiňuje některé možnosti, jak konstruovat výpočtově přijatelné modely, které nebudou narážet na bariéru výpočtové nezávládnutelnosti.

Text je zpracován přímo učebnicově pečlivě. Autor se neutápí ve formalismech, které by byly bez příslušného výkladu základních intuicí a motivací vedoucích k formulaci příslušných náročných teorií často (nejen pro „AI laika“) nesrozumitelné.

Takovýto přístup by však snadno mohl být na úkor přesnosti a vést až k populistickému způsobu výkladu. Myslím, že autorovi knihy se podařilo tomuto nebezpečí zcela vyhnout a nalézt opravdu vyvážený způsob, jak podat výklad srozumitelně a přitom teoreticky přesně. Je pochopitelné, že rozsah a šířka záběru způsobily, že některé partie nebylo možné dotáhnout do všech potřebných detailů, mnohá tvrzení jsou uváděna bez důkazu nebo je důkaz pouze naznačen. Na konci každé kapitoly najde čtenář komentář a shrnutí uvedené problematiky a zájemce o hlubší studium se může opřít o bohatou bibliografii. Text je v maximální míře soběstačný a neklade velké nároky na předpokládané vědomosti. Za velice šťastné však považuji zařazení tzv. dodatků (A – J) na konci knihy. Tyto dodatky plní více funkcí. Jednak pomohou správně pochopit hlavní text i čtenáři, který má jisté mezery v potřebných znalostech z oblasti logiky, teorie relačních databází, logického programování, algebry a teorie množin, teorie vyčíslitelnosti a složitosti, případně teorie algoritmů obecně. Tedy tyto pasáže nejsou v pravém slova smyslu dodatky, obsahují výklad nevyhnutelných předpokladů nutných k pochopení hlavního textu, a čtenář, který uvedenou problematiku neovládá, by je měl prostudovat spíše na začátku. Navíc je však možno v dodatcích nalézt formálnější výklad s přesnými definicemi, tyto části textu slouží jako formální doplněk určitých méně technických partií z jádra knihy.

Knihka může sloužit jako podklad pro minimálně semestrální přednášku o reprezentaci znalostí a metodách inference z hlediska AI na katedrách informatiky, tedy jako učebnice pro různé univerzitní kurzy umělé inteligence. Bude však jistě zajímavá pro všechny čtenáře, kteří touží získat nejnovější poznatky o studiu poznání a intelektu, tedy i pro ty, kteří se považují (stejně jako autorka této recenze) za „laiky“ v oblasti AI. Práci J. Šeřfránka lze považovat za velký přínos na slovenském a českém „trhu“. Podává srozumitelný a ucelený, přitom však teoreticky velice přesný sjednocující pohled na problematiku reprezentace znalostí, poznání a usuzování z perspektivy výpočtového modelu inteligence.

Marie Duží