

# Předmluva

Cílem této publikace je seznámit čtenáře se základními datovými strukturami a s metodami odhadu jejich složitosti. Význam návrhu vhodné reprezentace množin a hledání co nejefektivnějších algoritmů pro operace s nimi je dán jednak neustále se zvětšujícím objemem datových souborů a také tím, že nejzákladnější operace s daty (například vyhledávání) se v průběhu zpracování úloh nescíslněkrát opakují a i sebemenší zrychlení každé jednotlivé operace pak v praxi přináší nezanedbatelný zisk. Detailní teoretická analýza je přitom nezbytná, protože experimenty s velkými daty jsou časově náročné (k dosažení dostatečné vypovídací hodnoty se musí mnohokrát opakovat a je k tomu třeba zajistit si reprezentativní výběr testovacích dat) a jejich výsledky mohou být silně ovlivněny použitými programovacími technikami a konkrétním výpočetním prostředím. Nezanedbatelným problémem je i přesné změření času běhu samotného programu, zvláště když počítač prostřednictvím sítě sdílí více uživatelů pracujících na různých úlohách.

Teoretická analýza by na jedné straně měla být nezávislá na konkrétních podmínkách, na druhé straně by však její výsledky měly být použitelné v praktických situacích. Zvláště pro odhad doby výpočtu algoritmů je rozhodující volba modelu, ve kterém je budeme realizovat. Výpočetně jednodušší a na konkrétním počítači zcela nezávislý je model, ve kterém předpokládáme, že každá operace vyžaduje jednotku času, a doba výpočtu je pak dána celkovým počtem provedených operací (musí se však jednat o operace dostatečně elementární jako je např. dosazení hodnoty do proměnné nebo sečtení dvou čísel, nikoli už ale třeba sečtení většího počtu sčítanců nebo sečtení čísel se stovkami cifer). Typickým příkladem modelu tohoto typu je známý abstraktní model počítače RAM. Složitější, ale praxi více odpovídající, je model s váženými operacemi, kde váhy odrážejí relativní časovou náročnost instrukcí a musí být stanoveny podle parametrů konkrétního výpočetního prostředí (tj. počítače, operačního systému atd.). Při rozhodování, který model zvolíme, musíme tedy kromě obtížnosti výpočtů také uvážit, o čem by získané výsledky měly vypovídat a k čemu by měly sloužit. V každém případě je teoretická analýza sice pracnější než experimenty a je pro ni zapotřebí mít dostatečné znalosti z matematiky, především z kombinatoriky a teorie pravděpodobnosti, ale výsledky lze odvodit relativně rychle. Ideálním přístupem je oba způsoby analýzy vhodně kombinovat, tj. například prověřit teoreticky odvozené závěry o chování určitého algoritmu experimentem a na základě praktických zkušeností se pak pokusit o jeho eventuální vylepšení. V této knize se ale budeme věnovat výhradně teoretické analýze a experimentální výsledky (jsou-li známy) uvedeme pouze okrajově.

Knihla je členěna do pěti kapitol. V úvodní kapitole zavedeme základní pojmy a značení z teorie algoritmů, datových struktur a složitosti, které budeme v dalším textu používat. Také stručně vyložíme potřebný aparát z teorie grafů a teorie pravděpodobnosti.

Druhá kapitola bude věnována hašování, popisu několika různých metod řešení kolizí, univerzálnímu a perfektnímu hašování. Popíšeme i hašování pro externí paměti.

---

Ve třetí kapitole budeme řešit tzv. uspořádaný slovníkový problém nejprve pomocí uspořádaného pole a pak pomocí binárních vyhledávacích stromů. Velkou pozornost věnujeme vyváženým stromům – speciálně AVL-stromům, červeno-černým stromům a  $(a, b)$ -stromům.

Ve čtvrté kapitole budeme studovat několik nejpoužívanějších typů hald (regulární, leftist, binomiální, Fibonacciho) a obsahem páté kapitoly bude stručný přehled faktů o třídění. Budeme prezentovat nejpoužívanější algoritmy a metodu dolního odhadu složitosti třídění. Uvedeme zde také dva neznámější algoritmy pro hledání mediánu.

Dodatek obsahuje jednak některé složitější výpočty, jejichž zařazení přímo do textu by ubralo na přehlednosti, a dále pomocná tvrzení opakovaně používaná v různých kapitolách.

Pro větší přehlednost a snazší orientaci jsou všechny kapitoly členěny do sekcí a podsekcí, a lemmata, věty a důležité poznámky jsou číslovány pořadově v rámci kapitol. U algoritmů je kvůli lepší srozumitelnosti vždy nejprve uveden neformální slovní popis a pak zápis v tzv. pseudokódu, což je jakýsi mezistupeň mezi slovním vyjádřením a formálním zápisem v programovacím jazyce. Kvůli délce textu nebylo bohužel možné všechny algoritmy ilustrovat na příkladech a doplnit obrázky. Zařadili jsme proto jen ty nejdůležitější a zbytek přenecháváme čtenáři jako cvičení.

Každou kapitolu uzavírají bibliografické poznámky, které stručně shrnují historii dané problematiky a odkazují na některé publikace uvedené v závěrečném seznamu použité literatury. Tento seznam není a ani nemůže být vyčerpávající, protože publikací o datových strukturách jsou desetitisíce. Uvádíme v něm spíše klasické monografie a články a zájemce o současné výsledky odkazujeme na odborné časopisy. Ze starší literatury jsou to zejména monografie Knutha [K68, K73], Aho, Hopcrofta, Ullmana [AHU74] a Mehlhorna [M84]. Z novějších pak například tituly Gonnet a Baeza-Yates [GB91], Weiss [W92], Cormen, Leiserson, Rivest a Stein [CLRS01] nebo Mehlhorn a Sanders [MS08]. O teoretické analýze algoritmů pojednává kniha Sedgewicka a Flajoleta [SF96]. Sedgewickova monografie [S97] se zabývá datovými strukturami a algoritmy navíc i z hlediska implementace a programovacích technik.

Naše kniha je v první řadě určena jako učební text ke stejnojmenné přednášce na Matematicko-fyzikální fakultě UK, věříme však, že může posloužit i studentům informatiky jiných vysokých škol. Kromě znalosti středoškolské matematiky a elementárního programování (které je dnes rovněž standardní součástí středoškolských osnov) nevyžaduje žádné další předpoklady, protože všechny potřebné pojmy a vztahy jsou uvedeny přímo v textu. Přípravovaný druhý díl této publikace bude určen čtenářům s vyhraněnějším zájmem o tento obor, hlavně studentům vyšších ročníků magisterského studia a doktorandům. Představíme v něm další datové struktury (např. TRIE, samoupravující seznamy, splay-stromy), některé novější modifikace známých struktur (randomizované binární vyhledávací stromy, relaxované vyvažování, min-max haldy apod.) a další problémy (dynamizace, problém UNION-FIND).

---

Kniha vznikla na základě našich mnohaletých zkušeností s výukou tohoto předmětu, jehož náplň i forma postupně vykryštovaly do této současné podoby. Jsme si vědomi toho, že nároky na znalosti našich absolventů budou časem dále stoupat, a také, že tato publikace nepokryje vše, co od nich bude praxe vyžadovat. Doufáme ale, že jim poskytne solidní teoretický základ pro další studium i pro uplatnění v praxi.

Závěrem děkujeme našim kolegům z Matematicko-fyzikální fakulty UK za cenné názory a připomínky k výuce předmětu Datové struktury, a tím i k náplni této knihy. Studentům pak patří náš dík za připomínky k předběžné verzi tohoto textu, návrhy drobných vylepšení a upozornění na některé chyby, kterým jsme se přes veškerou snahu nebyli schopni vyhnout. Obzvláště ale děkujeme recenzentům za nesmírně pečlivou kontrolu rukopisu. Jejich posudky totiž kromě výčtu konkrétních nedostatků obsahovaly řadu námětů na vylepšení textu jak po formální, tak po obsahové stránce, a upozornily nás i na některé související problémy, které by bylo vhodné zmínit. Ne všechny jsme do tohoto textu zapracovali, protože jeho rozsah by se tím neúměrně zvětšil, budeme ale uvažovat o jejich eventuálním zařazení do druhého dílu této publikace.

Praha, červen 2010

Autoři