

**BAEZA-YATES, Ricardo,
RIBEIRO-NETO, Berthier.
Modern information retrieval.
New York : Addison-Wesley, 1999.
513 s.
ISBN 0-201-39829-X.**

„Jaká je dobrá kniha o vyhledávání informací?“ To je otázka, kterou lze slyšet velmi často, a to nejen mezi studenty zaměřenými na informatiku, popř. informační vědu a práci s informacemi obecně. Odpověď nikdy nebyla jednoznačná. Dlouhá léta byl výběr omezen na knihy, na nichž se podílel Gerard Salton. Jednalo se především o tyto publikace:

- SALTON, Gerard, MCGILL, Michael J. *Introduction to modern information retrieval.* New York : McGraw-Hill, 1983. 448 s.
- SALTON, Gerard. *Automatic text processing : the transformation, analysis, and retrieval information by computer.* Reading (MA) : Addison-Wesley, 1988. 530 s.

Doporučovanou „klasikou“ byla také kniha od Van Rijsbergena z roku 1979 (VAN RIJSBERGEN, C. J. *Information Retrieval.* London : Butterworths, 1979. 208 s. Dostupná též na: <http://www.dcs.gla.ac.uk/Keith/Preface.html>).

V 90. letech 20. století došlo k nárůstu zájmu o problematiku vyhledávání informací jednak ze strany uživatelů, jednak z hlediska vývoje potřebného softwaru. Důvody jsou zřejmé – nastoupený trend dostupnosti informací v elektronické primární podobě, který pak byl ve 2. polovině 90. let ještě zvýrazněn expanzí internetu a služby WWW. Jakkoli lze považovat výše uvedené monografie za nadčasové, bylo zapotřebí zohlednit nové vlivy a nové výzvy, a proto vznikla recenzovaná kniha, která se snaží zmíněnou mezeru zaplnit.

Kniha vyčerpávajícím způsobem předestírá problematiku vyhledávání informací z počítačového hlediska. Jsou zde uvedeny algoritmy, vzorce, principy fungování různých informačně-vyhledávacích modelů, dotazovacích jazyků, indexů, uživatelského rozhraní a vizualizace informací. Devět kapitol je dílem obou hlavních autorů, o šest se pak podělili další přední odborníci.

Knihu lze použít jako učební text pro různé úrovně čtenářů. V úvodu jsou navržena variantní schémata čtení jednotlivých kapitol. Uvedená schémata ostatně odůvodňují pro někoho možná netradiční řazení kapitol, ve kterém je např. kapitola věnovaná hodnocení úspěšnosti vyhledávání zařazena již před problematiku dotazovacích jazyků a indexování. Každá kapitola obsahuje v závěru bibliografický komentář, který doporučuje ke studiu další relevantní publikace. Samotný seznam pramenů pro celou knihu je úctyhodný – zahrnuje 852 položek. Podrobnější představu o obsahu knihy si lze udělat na základě stránky <http://sunsite.dcc.uchile.cl/irbook/>, kde je mj. zpřístupněn úplný obsah, předmluva, dvě vybrané kapitoly (Úvod a kap. 10 – Uživatelské rozhraní a vizualizace) a slovník pojmů.

Kniha je užitečná nejen pro svůj téměř encyklopedický přehled, ale i při hlubším zájmu o některý ze specifických

kých problémů vyhledávání informací. Za úspěch lze považovat, že se autorům podařilo zohlednit vlivy internetu prakticky ve všech aspektech vyhledávání informací. Autoři považují za samozřejmé, že internet narušil některá paradigmatata. Celá řada tradičních principů je modifikována: práce s kořeny slov (stemming) či vylučování stopslov poněkud ustoupila do pozadí (nicméně to platilo v době vydání knihy, nyní je opět trochu jiná situace), vzniká poptávka pro strukturovaném vyhledávání, které narušuje obvyklé vnímání textu v lineární podobě apod.

Úvodní kapitola vymezuje oblast vyhledávání informací včetně krátké diskuse o vlivu internetu. Kapitola 2 představuje různé modely používané pro vyhledávání informací – booleovský, vektorový, pravděpodobnostní – včetně jejich alternativ či rozšíření (fuzzy množiny, neuronové sítě, bayesovské sítě apod.). Kapitola 3 probírá možnosti hodnocení efektivity vyhledávání (úplnost, přesnost) včetně alternativních přístupů, zmiňuje i testovací kolekce dokumentů. Kapitola 4 je věnována dotazovacímu jazykům, kapitola 5 pak zpracování dotazů. Kapitola 6 ukazuje možnosti vyhledávání ve smíšených dokumentech, přičemž je zpracována problematika metadat, značkovacích jazyků a různých multimediálních formátů. Kapitola 7 je zaměřena na specifické textové operace (stemming, stopslova, výběr indexovaných termínů, role tezaurů, shlukování dokumentů, komprese textu). Kapitola 8 podrobně popisuje datové struktury a algoritmy pro vyhledávání (invertované soubory, jiné textové indexy, sekvenční vyhledávání, porovnávání vzorů, strukturované dotazy aj.). Kapitola 9 je úvodem do problematiky paralelního a distribuovaného vyhledávání. Cílem

kapitoly 10 je přiblížit vliv uživatelského rozhraní a vizualizace výsledků vyhledávání na efektivnost práce uživatele. Kapitoly 11 a 12 rozpracovávají specifické problémy multimediálních informačních systémů. Kapitola 13 je věnována vyhledávání informací v prostředí webu, přičemž jsou uvedeny principy fungování základních typů vyhledávacích služeb (vyhledávací stroje, metavyhledávací služby, katalogy), rozebrán je také vliv hypertextových odkazů. Kapitola 14 pokrývá tradiční bibliografické systémy. Kapitola 15 seznamuje čtenáře s principy digitálních knihoven. V samostatné příloze je detailně popsán Porterův algoritmus pro stemming, součástí knihy je rovněž slovník pojmů.

Přestože jsou jednotlivé kapitoly napsány velmi fundovaně, nepodařilo se vzhledem k počtu spoluautorů vyhnout určitým rozporům. Například v jedné kapitole jsou konstatovány fundamentální lexikální problémy při zpracování textu, zatímco jinde je prezentována technika, která vychází z předpokladu, že zpracování textu je triviální.

Kniha sice vyšla již před téměř pěti lety, ale stále má co říci; řada poznatků je nadčasová nebo je platná i dnes – jsou zde např. popisovány principy algoritmu Page-Rank (Google), který v době vydání knihy teprve začínal dobývat svět webových vyhledávacích strojů. Kromě toho od té doby žádná jiná monografie s podobnou šíří záběru nevyšla. Je pravda, že se od roku 1999 objevila řada významných podnětů, jako je např. vyhledávání v XML dokumentech, sémantický web, využití ontologií apod., ale to nic neubírá na konstatování, že knihu lze stále považovat za solidní odrazový můstek pro proniknutí do principů vyhledávání informací.

Vilém Sklenák
Katedra informačního
a znalostního inženýrství,
fakulta informatiky a statistiky,
Vysoká škola ekonomická

