

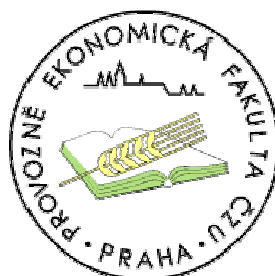
ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

PROVOZNĚ EKONOMICKÁ FAKULTA



Vyhledávání informací na Internetu

Teze diplomové práce



Vedoucí diplomové práce:

Ing. Hana RYSOVÁ

Autor :

Bc. Jan NEHASIL © 2005

Tato práce se zabývá problematikou vyhledávání informací na Internetu. Cílem je popsat a zhodnotit Internet jako neobsáhlejší informační zdroj, jeho vznik, vývoj a současné trendy ve vyhledávání informací pomocí tohoto informačního média.

V první kapitole mé práce jsou definovány základní pojmy. Informací chápeme údaj o reálném prostředí, o jeho stavu a procesech v určitém čase. Vyhledáváním se v této práci rozumí použití určitého vyhledávacího systému k nalezení výsledku. Internet je celosvětová počítačová síť, původně sloužící k armádním a vědeckým účelům, dnes největší informační zdroj a součást každodenního života moderní vyspělé společnosti.

Internet a potřeba v něm rychle a efektivně bez velkých znalostí vyhledávat, daly vzniknout relativně nové technologii sběru dat. První zdařené pokusy vyhledávacích systémů se objevují v polovině devadesátých let ve Spojených státech amerických. Původně vznikly pro usnadnění práce pouze svým vlastním tvůrcům, většinou na akademické půdě. Ovšem s rychlým přiblížením se Internetu běžným uživatelům byly zpřístupněny veřejnosti a jsou dnes hojně užívanou službou poskytovanou na Internetu.

Právě popisem těchto služeb a jejich užití ve vyhledávacích systémech se zabývá velká část mé práce. Aby to bylo možné, bylo nejdříve nutné rozdělit vyhledávače do několika základních skupin. Na základě tohoto rozdělení byly jednotlivé skupiny teoreticky popsány včetně jejich znaků a specifikací. Jako první jsou katalogové servery, lidmi manuálně vytvářené seznamy odkazů na webové stránky uspořádané do hierarchické struktury. Mají na své titulní stránce hlavní kategorie (cestování, kultura, obchod, počítače, zábava apod.), které se případně větví na podkategorie. Uživatel tak nejen vidí některé podkategorie hlavní větve, ale navíc si je může ihned prohlédnout. Další skupinou jsou vyhledávací servery tzv. fulltexty. Ty oproti manuálně tvořeným katalogovým serverům využívají automatické mechanismy k vytvoření vlastní databáze klíčových slov z jednotlivých stránek. Nejen, že jsou tyto databáze až tisíckrát větší než u katalogových serverů, ale vyhledávací servery nabízejí ještě navíc možnosti přesnější specifikace dotazu pomocí různých operátorů např. AND, OR atd. Za třetí skupinu lze považovat souhrnné vyhledávací servery a metavyhledávače. V této práci jsou popsány jako samostatná skupina, přestože jejich princip spočívá ve využití předcházejících katalogů a vyhledávacích serverů. Poslední skupinou patřící také mezi vyhledávače jsou PPC vyhledávače. Pay-per-click (PPC) vyhledávače fungují jako klasické vyhledávací

servery s tím rozdílem, že za určité velmi frekventované termíny se platí. Nezáleží tedy na tom, jak je stránka optimalizována pro web, ale na tom, kolik jste ochotni zaplatit za každého návštěvníka. U každé z výše zmiňovaných skupin vyhledávačů jsou uvedeny příklady jejich zástupců.

Internet však nabízí celé spektrum dalších informačních nebo zpravodajských zdrojů, které jsou velmi často také přístupné na webu, ale které pomocí výše uvedených vyhledávacích služeb uživatel nenajde. Jsou to zejména hudební nahrávky, software všeho druhu, elektronické adresy nebo firemní, průmyslová a obchodní data. Součástí této práce je i krátký rozhovor s ministrem informatiky Vladimírem Mlynářem na téma Portálu veřejné správy.

Hlavní a stěžejní částí této práce bylo provést důkladný test dnešních vyhledávačů. Autor vybral 10 vyhledávačů, které reprezentují různé vyhledávací skupiny. Vybraným vyhledávačům bylo položeno 20 unikátních dotazů. Vzhledem k absenci podobně rozsáhlého testu v ČR je nutno poznamenat, že autor neměl příliš mnoho pramenů k porovnání. Autor práce konzultoval metodiku testu s předními odborníky na tuto problematiku a vycházel i z poznatků a závěrů jejich prací. K hodnocení vyhledávačů byl využit bodovací systém navržený právě po těchto konzultacích. Ze závěrů je patrné, že nejlepší výsledek v dnešní době poskytují vyhledávací služby kombinující kvalitní katalog a prověřený fulltextový vyhledávač. Své výsadní postavení na poli vyhledávačů ukázal americký produkt Google, který dopomohl českému portálu Tiscali k vítězství v testu. Na velmi dobrém třetím místě se umístil portál Seznam využívající Jyxo. Výsledky testu byly velmi vyrovnané v první osmičce, jen na posledních místech s velkou ztrátou dopadly vyhledávače Red Box a Yo. Doplnkem byl test zaměřený na složení úvodních stran českých vyhledávačů. Potvrdila se tvrzení odborníků o přílišném objemu úvodních stran některých vyhledávačů. S rostoucím objemem, zejména grafických prvků, roste i čas jejich stažení na straně uživatele. V tomto testu obstály nejlépe čistě fulltextové vyhledávače Jyxo a Google s minimem grafiky a maximální efektivitou práce. Nejhuře dopadly portály Volný, Centrum a Yo, které jsou velmi graficky objemné.

Výhledem do budoucna se zabývá předposlední kapitola. V budoucnu bude snaha provozovatelů vyhledávacích služeb co nejvíce přiblížit Internet běžným uživatelům a maximálně zefektivnit svoje služby. Některé tendence jsou vidět již dnes.

Objevil se první mluvící vyhledávač Speegle, který nejen vyhledává, ale i přečte, co našel. Také je tu inteligentní vyhledávač Brainboost, který odpovídá na prostě položené otázky. Výzvou do budoucna je i sémantický web, který stále naráží na neochotu tvůrců stránek upravit své weby do té správné podoby. Už dnes můžeme ale říci, že ať už se vyhledávací služby Internetu zlepšují jakkoliv, nikdy nejspíše nebudou schopny poskytnout jediný výsledek přesně odpovídající dotazu či požadavku uživatele. Přesto je zřejmé, že lepší technologie jsou zárukou a příslibem pro generování přesnějších výsledků vyhledávání, v konečném důsledku doprovázeného zvýšením loajality uživatelů i zákazníků internetových firem.

Internet je velmi proměnlivé médium. Platí tu, že co najdeme dnes, nemusí již zítra existovat a naopak. Proto umění rychle a efektivně v něm vyhledávat by mělo patřit mezi základní dovednosti moderního uživatele. Kdo toto umění neovládá, nikdy plně nevyužije potenciál a možnosti, jenž Internet přináší.

Seznam použitých zdrojů literatury

Knižní zdroje

- [1] BRÁZA, Jiří. *Internet.cz jak hledat a najít*. Praha: Grada Publishing, 2001. 128 s. ISBN 80-247-9002-5.
- [2] NONDEK, L.a ŘENČOVA, L. *Internet a jeho komerční využití*. Praha: Grada Publishing, 2000. 117 s. ISBN 80-7169-933-0.
- [3] HLAVENKA, Jiří. *Mistrovství ve vyhledávání na Internetu*. Praha: Computer Press, 2002. 195 s. ISBN 80-7226-759-0.
- [4] VOŘECH, J.a MORKES D. *1001 tipů a triků pro Internet*. Brno: Computer Press, 1999. 393 s. ISBN 80-7226-177-0.
- [5] HAVLÍČEK, Zdeněk. *Informační technologie v řízení*. Praha: ČZU, 2000. 81 s. ISBN 80-213-0688-2.
- [6] PETERKA, Jiří; ČERMÁK, Miloš a MATOUŠEK, Petr. *Se z@vináčem na INTERNET*. Praha: Academia, 1999. 309 s. ISBN 80-200-0787-3.
- [7] NEMRAVA, Jan. *Optimalizace www stránek pro vyhledávací a indexování katalogy*. VŠE Praha, 2004. 69 s.

Časopisecké zdroje

- [8] DVOŘÁČEK, Martin. *Kdo hledá, najde?*. Chip, 2001, roč. 11, č. 10, s 110-113.
- [9] BITTER, Ladislav. *Jak se hledá na českých portálech*. Chip, 2002, roč. 12, č. 6, s 87-92.
- [10] CHURÁ, M.a VOSTRÝ, P. *Pohled pod pokličku českých portálů*. Chip, 2002, roč. 12, č.1, s 78-81.

Internetové zdroje

- [11] Portál Centrum
<http://www.seznam.cz/>.
- [12] MIŽOCH, Lukáš. Open directory project (OPD). *Lupa* [online]. 2002, [cit. 12. února 2005]. URL: <<http://www.lupa.cz/clanek.php3?show=2272>>.
- [13] NEHASIL, Jan. Registrujte stránky do českých vyhledávačů. *Technet* [online]. 2005, [cit. 12. února 2005]. URL: < <http://technet.idnes.cz>>.

- [14] PROKOP, Marek. PPC vyhledávače na českém Internetu. *Lupa* [online]. 2002, [cit. 12. února 2005]. URL: <<http://www.lupa.cz/clanek.php3?show=2602>>.
- [15] PCC vyhledávač eTarget
<http://www.etaget.cz/>
- [16] NEHASIL, Jan. Ticho, vyhledávače mluví!. *Lupa* [online]. 2005, [cit. 22. února 2005]. URL: <<http://www.lupa.cz/clanek.php3?show=3978>>.
- [17] NEHASIL, Jan. Velké problémy neznámého portálu. *Lupa* [online]. 2005, [cit. 22. března 2005]. URL: <<http://www.lupa.cz/clanek.php3?show=4037>>.
- [18] BERNERS-LE, Tim. HENDLER, James. LASSILA, Ora. The semantic web. *Scientific American* [online]. 2001, [cit. 15. února 2005]. URL: <http://www.sciam.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>.
- [19] BRAI, Tim. PAOLI, Jean. MALER, Eve. Extensible Markup Language (XML) 1.0 (Third Edition) [online] 2004, [citace 12. února 2005]. URL: <<http://www.w3.org/TR/REC-xml>>.