

# Problémy súčasného internetu a navrhované riešenia

Doc. PhDr. Soňa Makulová, Ph.D.,

Katedra knižničnej a informačnej vedy, Filozofická fakulta Univerzity Komenského Bratislava,

Gondova 2, 818 01 Bratislava Slovensko, e-mail: [makulova@elt.sk](mailto:makulova@elt.sk)

[http://www.fphil.uniba.sk/~kkvi/kto\\_je\\_kto/makulova.html](http://www.fphil.uniba.sk/~kkvi/kto_je_kto/makulova.html)

## Úvod

O tom, že internet už má svoje pevné miesto v našom každodennom živote dnes už nikto nepochybuje. Internet môžeme definovať ako informačné, komunikačné, marketingové, reklamné a obchodné médium, ktoré využívalo podľa štatistík Nua v septembri 2002 takmer 606 miliónov používateľov na celom svete.

Počet internetových používateľov narastá aj na Slovensku. Podľa prieskumov spoločnosti Taylor Nelson Sofres internet využívalo v decembri 2001 viac ako 23% populácie starších ako 18 rokov v porovnaní s 5% populácie v roku 1997. Napriek tomu je percento populácie využívajúce internet na Slovensku ešte stále veľmi nízke. V januári 2003 University of California - Los Angeles (UCLA) publikovala prehľad o využívaní internetu americkou populáciou. Podľa dokumentu až 71% americkej populácie využívalo internet v roku 2002 (The UCLA 2003).

V našom príspevku sa zameriame predovšetkým na internet ako informačné médium a na problémy, pred ktorými stoja používatelia pri vyhľadávaní informácií v internete. Je to veľmi dôležité, pretože podľa výskumov University of California množstvo digitálnych informácií neuveriteľne narastá. V štúdií Lymana a Variana "How much information?" (2000) sa uvádza, že zatiaľ čo každý deň vychádza vo svete v tlačenej forme približne jeden milión kníh, v elektronickej forme je to viac ako 7 miliónov. Internet obsahoval v roku 2000 vyše 550 miliárd dokumentov. Čo je ale neuveriteľné, tlačенý obsah reprezentuje iba 0,003% celého obsahu publikovaného vo svete. Znamená to, že každej jednej vete v tlačných médiách zodpovedá 30 000 viet v digitálnej forme.

V súčasnosti prebiehajú mnohé prieskumy zaoberajúce sa skúmaním správania sa používateľov pri vyhľadávaní informácií v internete. Zatiaľ čo vyhľadávanie v komerčných databázových centrách sa väčšinou uskutočňovalo prostredníctvom informačných profesionálov, v internete sa stále viac používateľov spolieha na vlastné sily. Pritom výskumy jasne dokazujú, že iba málo používateľov rozumie tomu, ako vyhľadávacie nástroj interpretuje požiadavku, aké sú rozdiely medzi vyhľadávacími nástrojmi v internete a v databázových centrách, akým spôsobom sa používajú operátory Boolovej algebry a pod.

V septembri a októbri 2002 sme realizovali s podporou spoločnosti EL&T, mediálnou agentúrou SITA, vydavateľstvom Computer Press a spoločnosťou EuroTel rozsiahly prieskum zameraný na využívanie informačných zdrojov internetu a miery používateľskej spokojnosti s vyhľadávacími nástrojmi formou dotazníka. Prieskum bol podporovaný reklamnými bannermi nachádzajúcimi sa na navštevovaných www serveroch a motiváciou pre vyplnenie dotazníka boli aj zaujímavé ceny, ktoré ponúkli sponzori súťaže. Prieskumu sa zúčastnilo 1676 respondentov, ktorí odpovedali na 40 otázok. Otázky sa týkali používaných technológií a spôsobu pripájania, používaných vyhľadávacích nástrojov a samotného obsahu. Išlo o prvý prieskum tohoto druhu na Slovensku. O výsledkoch prieskumu sa zmienime pri jednotlivých problémových okruhoch.

## Informačné preťaženie a nájdenie relevantnej informácie ako najväznejší problém súčasného internetu

Problémy súčasného internetu by sme mohli zhrnúť nasledovne:

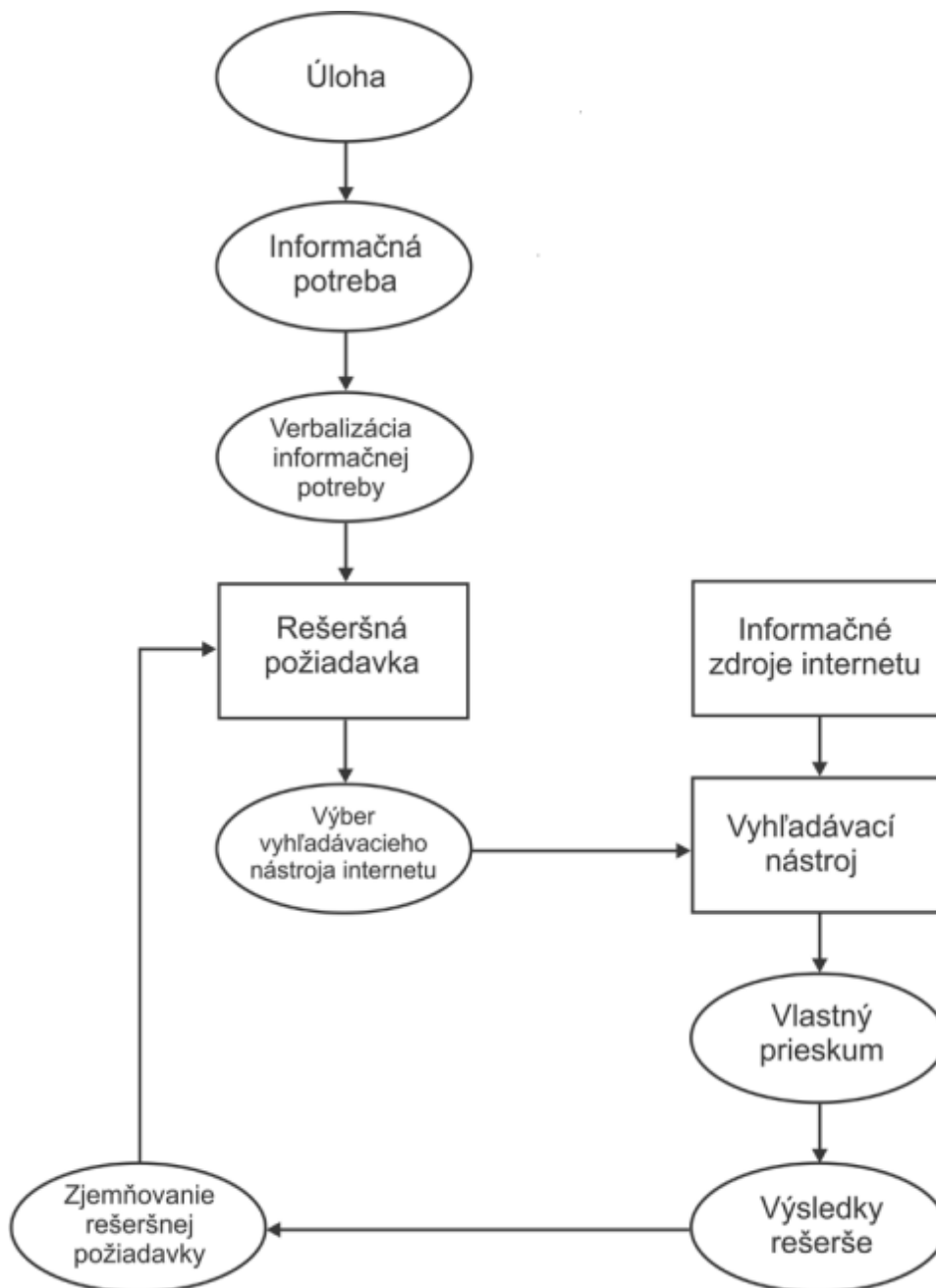
- neustály nárast hosťiteľských počítačov internetu, informačných kanálov a informačných zdrojov,
- nové typy dokumentov v prostredí internetu (multimédiá, virtuálne hypertextové dokumenty, neustála aktualizácia dokumentov, rozdielne informačné jednotky), ktoré sú nedostatočne indexované, respektíve nie sú indexované vôbec,
- množstvo vyhľadávacích nástrojov rôznej kvality s rôznou bázou dát, s nedostatočnou používateľskou podporou,
- žiaden vyhľadávací nástroj nepokrýva celý web a preto, ak kladieme vysokú požiadavku na úplnosť, musíme využívať množstvo rôznych vyhľadávacích nástrojov, pričom nie je k dispozícii dokumentácia o metódach indexovania, vyhľadávania, šírke pokrytia z rôznych hľadísk (typy informačných prameňov, jazykové hľadisko, geografické hľadisko, predmetové hľadisko a pod.),
- obsah a lokalizácia dokumentov sa často menia, nie je jasná frekvencia aktualizácie indexov databáz, stále viac je aktuálny problém manažovania dokumentov v prostredí webu,
- nedostatok inteligentných nástrojov na vyhľadávanie,
- vyhľadávacie nástroje sa väčšinou sústreďujú na úplnosť rešerše, to znamená, že je snaha indexovať čo najväčšiu časť webovského priestoru, iba málo nástrojov umožňuje zlepšiť presnosť vo vyhľadávaní (spätná väzba, využívanie slovníka, pojmové vyhľadávanie),
- nedostatočná používateľská podpora a podpora na využívanie rozšíreného vyhľadávania,
- iba málo vyhľadávacích nástrojov hodnotí kvalitu informačných zdrojov a využíva niektorý systém klasifikácie na ich triedenie,
- súčasná verzia jazyka HTML neumožňuje dostatočne popísať obsah informácie (metainformáciu) a informačnom prameni.

Pri analýze problémov, s ktorými sa stretáva používateľ pri vyhľadávaní informácií v internete budeme vychádzať zo zjednodušeného modelu informačného prieskumu (obr. 1).

Uvedený model informačného prieskumu zodpovedá prieskumovému procesu v internete pozostávajúcemu z nasledovných krokov:

1. Formulácia a analýza rešeršnej požiadavky
2. Výber vhodného vyhľadávacieho nástroja internetu
3. Kombinácia selekčných termínov pri vyhľadávaní
4. Vlastný prieskum
5. Zjemňovanie a ladenie rešeršnej požiadavky

Na základe uvedeného modelu informačného prieskumu sa pokúsime analyzovať súčasné problémy pri vyhľadávaní a načrtnúť niektoré riešenia.



Obr. 1 Model informačného prieskumu v internete

### Problémy pri formulácii a analýze rešeršnej požiadavky

Formulácia a analýza rešeršnej požiadavky patria k najdôležitejším krokom prieskumového procesu. Vzhľadom na množstvo a zložitosť vyhľadávacích nástrojov, ale aj zlé výsledky, ktoré dosahujú používatelia pri vyhľadávaní informácií, treba venovať väčšiu pozornosť analýze správania sa používateľa pri informačnom prieskume. Na základe mnohých výskumov týkajúcich sa vyhľadávania informácií bolo zistené že (Makulová, 2002):

- väčšina požiadaviek má v priemere 2,35 termínov,
- jednotlivé termíny nevystihujú obsah toho, čo chce používateľ nájsť,
- používatelia nevedia používať Boolovské operátory,
- 85% používateľov si pozrie iba nájdené záznamy na jednej obrazovke,

- 78% požiadaviek sa ďalej nemodifikuje,
- 22% používateľov hľadá určitú stránku,
- 24% používateľov si chce z internetu niečo stiahnuť,
- 10% používateľov na internete nakupuje,
- veľa používateľov nevyhľadáva, ale iba sleduje spojenia.

Pritom používatelia chcú a hľadajú informácie na internete. V nasledujúcej tabuľke vidíme, koľko prieskumov sa vykoná denne v najznámejších vyhľadávacích nástrojoch (Sullivan, 2003).

Nástroj	Počet prieskumov za deň	Poznámka
Google	250 miliónov 130 miliónov	február 2003 október 2001 (Polovica je na Google.com a zvyšok na partnerských sídlach)
Overture	167 miliónov	február 2003
Inktomi	80 miliónov	február 2003
LookSmart	45 miliónov	február 2003
FindWhat	33 miliónov	január 2003
Ask Jeeves	20 miliónov	február 2003
AltaVista	18 miliónov	február 2003
FAST	12 miliónov	február 2003

### Tabuľka č. 1 Počet denných prieskumov v najznámejších vyhľadávacích nástrojoch

O nedostatočnej informačnej gramotnosti svedčí aj spôsob, akým používatelia formulujú svoje rešeršné požiadavky. Chris Sherman (2002) uverejnil v svojom spravodaji SearchDay najviac vyhľadávané termíny za rok 2001 vo vyhľadávacích nástrojoch Google, Yahoo a Lycos. V prípade Lycosu je v zátvorke porovnané umiestnenie termínu za rok 2000.

Google	Yahoo	Lycos
1. nostradamus 2. cnn 3. world trade center 4. harry potter 5. anthrax 6. windows xp 7. osama bin laden 8. audiogalaxy 9. taliban 10. loft story	1. PlayStation 2 2. Britney Spears 3. WWF 4. Dragon Ball Z 5. Napster 6. World Trade Center 7. Harry Potter and the Sorcerer's Stone 8. Dale Earnhardt 9. NASCAR 10. Internal Revenue Service	1. Dragonball (2) 2. Britney Spears (1) 3. Napster (8) 4. Tattoos (7) 5. Osama bin Laden (-) 6. IRS (23) 7. Pokemon (3) 8. World Trade Center (-) 9. Nostradamus (-) 10. WWF (4)

### Tabuľka č. 2 Najviac vyhľadávané termíny v niektorých svetových vyhľadávačoch za rok 2001

Ako ľudia formulujú svoje informačné potreby a čo ich zaujíma v určitom časovom období napovedá aj služba Lycos 50, ktorá monitoruje najviac používané vyhľadávacie termíny. Sieť Lycos navštívi každý deň okolo 12 miliónov používateľov, pričom jednotlivé dotazy majú priemerne 2,3 termínov. Tieto predstavujú základ týždenných správ. V čase písania článku kedy

vrcholila vojna v Iraku bola najvyhľadavanejším termínom minister informácií Iraku Mohammed Saeed al-Sahaf. Služba je prístupná na adrese <http://50.lycos.com>.

Na Slovensku ľudia za rok 2001 na portáloch ATLAS.SK a Centrum najviac hľadalo slovo SMS. V prvej desiatke ďalej dominovalo slovo mobil, témy zábavy a uvoľnenie (zoznamka, chat, video atď.) a zamestnania (práca, zamestnanie, brigáda).

V nasledujúcej tabuľke sú uvedené najviac vyhľadávané slová na portáloch ATLAS.SK, Centrum a SURF.SK.

ATLAS.SK	Centrum	SURF.SK
SMS	SMS	sex
mp3	sex	SMS
sex	www	gay
porno	porno	mp3
mobil	praca	erotika
mapy	mob	porno
praca	CD	lolita
pohľadnice	mobil	praca
zoznamka	video	pohľadnice
gay	chat	inzercia
erotika	web	video
inzercia	zamestnanie	nokia
hry	zoznamka	mobil
logo	brigada	reality
nokia	obchod	knihy
chat	zoznamenie	chat
práca	pokec	download
auto	globtel	obrázky
referaty	sutaz	hry
zamestnanie	shop	zadarmo

### Tabuľka č. 3 Najviac vyhľadávané termíny na niektorých slovenských vyhľadávačoch za rok 2001

V mesiacoch január až marec 2003 sme analyzovali transakčné logy na serveri SURF.SK. (<http://www.surf.sk>), ktorý obsahuje informácie o viac ako 20 000 webových stránkach slovenského internetu. Adresár je hodnotený, najlepšie stránky sú v kategórii TOP. Systém má veľmi dobre prepracované rozšírené vyhľadávanie. Napriek tomu pri analýze 500 najviac používaných dotazov zo 40 000 sme zistili, že iba niekoľko požiadaviek bolo formulovaných ako viacslovné spojenie (napaľovanie CD, erotické povídky, cestovná kancelária, práca v zahraničí, teen sex, tetris 4000, logo zadarmo, britney spears, Harry Potter, ponuka práce, obchodný register, free SMS, Banská Bystrica, brigády v zahraničí, pracovné príležitosti).

K najviac využívaným termínom patrili slová sex (frekvencia výskytu 660), sms (494), gay (298), pohľadnice (277), mp3 (264), erotika (236), porno (156), incest (152), práca (143) a pod.

Je to dané pravdepodobne tým, že iba veľmi málo používateľov slovenského internetu využíva zdokonalené vyhľadávanie. Príčinou je nedostatočný používateľský interfejs, žiaden z najznámejších slovenských vyhľadávacích nástrojov neponúka na formuláciu rozšíreného

vyhľadávania preddefinovaný formulár. Podobne sa nedodžiavajú odporúčania informačných architektov týkajúce sa vyhľadávania na webových stránkach.

Zarážajúcou skutočnosťou je, že najnavštevovanejší slovenský portál Zoznam slovenského Internetu neposkytuje možnosť rozšíreného vyhľadávania a ani súbory typu help, pritom má podľa výsledkov auditu spoločnosti Taylor Nelson Sofres pravidelne viac ako 36 000 jedinečných návštevníkov denne. V januári 2003 ho navštívilo 360 776 jedinečných návštevníkov, čo predstavovalo takmer 41,49% internetovej populácie na Slovensku.

Jansen, Spink a Saracevic (2000) skúmali transakčné logy 51 473 dotazov 18 113 používateľov vyhľadávacieho portálu Excite. Na základe analýzy konštatovali, že vyhľadavanie na internete prebieha iným spôsobom ako v tradičných databázových centrách, na CD ROMoch, dokonca aj v online katalógoch knižníc. Používatelia málo používajú zdokonalené vyhľadavanie, iba málo používateľov ďalej modifikuje rešeršnú požiadavku a priemerná požiadavka obsahuje 2,21 termínov. Približne iba 8,5% dotazov obsahovalo Boolovské operátory a iba 9% obsahovalo niektorý iný operátor zdokonaleného vyhľadávania.

Používanie Boolovských operátorov v internete je oveľa nižšie ako v databázových centrách. Siegfried, Bates a Wilde (1993) uskutočnili výskum, kde skúmali, akým spôsobom vyhľadávajú vedci z oblasti humanitných vied v DIALOGu. Boolovské operátory využívalo viac ako 36% používateľov.

Jansen (2000) skúmal ďalej, akým spôsobom ovplyvní štruktúra rešeršnej požiadavky využívajúca možnosti zdokonaleného vyhľadávania výsledky rešerše. Výskum uskutočnil vo vyhľadávačoch AltaVista, Fast, Excite, Northern Light a Infoseek. Využil 75 jednoduchých a 150 zložitých požiadaviek, pričom z každej sa skúmalo iba prvých 10 nájdených záznamov. Každá z jednoduchých požiadaviek sa ďalej modifikovala s využitím operátorov Boolovej algebry. Pri analýze nájdených 2768 výsledkov sa zistilo, že rozdiel pri jednoduchom vyhľadávaní a zdokonalenom vyhľadávaní bol iba v 2,7 rozdielnych záznamoch. Znamená to, že v priemere až 7,3 záznamov, ktoré sa našli pri jednoduchom vyhľadávaní boli aj vo výsledkoch rešerše pri zložitom vyhľadávaní. Je potom otázne, či má význam využívať zdokonalené vyhľadavanie iba pri tak malej odchýlke. Navyše používanie Boolovských operátorov zvyšuje riziko chybovosti. Riešenie vidíme predovšetkým v návrhu používateľského interfejsu, ktorý minimalizuje chybovosť a umožňuje ďalej modifikovať a meniť rešeršnú požiadavku.

To zároveň vysvetľuje, prečo je veľa používateľov spokojných iba s výsledkami pri jednoduchom vyhľadávaní a svoju požiadavku ďalej nemodifikuje. Podobne sa dá predpokladať, že algoritmus radenia nájdených záznamov vychádza zo správania sa používateľov a ako prvé sú umiestnené tie záznamy, ktoré obsahujú všetky hľadané termíny a hľadané termíny sa vyskytujú vedľa seba.

## Výsledky prieskumu správania sa slovenských používateľov pri vyhľadávaní informácií

V septembri a októbri 2002 sme realizovali s podporou spoločnosti EL&T, mediálnou agnetúrou SITA, vydavateľstvom Computer Press a spoločnosťou EuroTel rozsiahly prieskum zameraný na využívanie informačných zdrojov internetu a miery používateľskej spokojnosti s vyhľadávacími nástrojmi formou dotazníka. Prieskum bol podporovaný reklamnými príškami nachádzajúcimi sa na navštevovaných www serveroch a motiváciou na vyplnenie dotazníka boli aj zaujímavé ceny, ktoré ponúkli sponzori súťaže. Podrobné výsledky prieskumu ako aj otázky sú uvedené v štúdií Informačné správanie používateľov pri vyhľadávaní informácií v internete (Makulová, 2003).

Naša hypotéza, že používatelia slovenského internetu nepoznajú princíp práce a rozdiel medzi rôznymi vyhľadávacími nástrojmi sa potvrdila. Až 51,19% používateľov nevie, aký je rozdiel medzi predmetovým adresárom, prieskumovým strojom a metaprieskumovým strojom a iba

29,77% to vie čiastočne. Skoro rovnaký počet 43% používateľov pozná aj predmetové adresáre aj prieskumové stroje, ale iba 18,97% používateľov pozná metaprieskumové stroje.

Čo je zaujímavé, až 69,93% používateľov je vo väčšine prípadov spokojných s nájdenou informáciou a v polovici prípadov je to 19,99%. Svedčí to o tom, že používatelia nedokážu ešte stále hodnotiť informačné zdroje a zabúdajú na to, že kvantita ešte neznamená kvalitu.

Až 81,8% používateľov využíva najčastejšie Google, nasleduje Yahoo (44,63%) a AltaVista (43,2%). Zo slovenských je to už tradične Zoznam slovenského Internetu (79,36%), Google.sk (53,1%), Superzoznam (39,62%), ATLAS.sk (25,24%) a Centrum.sk (20,53%).

Zaujímavým zistením je, že až 34% používateľov udáva, že využíva v rovnakej miere aj slovenské aj svetové vyhľadávacie služby. Znamená to, že internet sa stáva významným zdrojom informácií, o čom svedčí aj skutočnosť, že až 78,16% respondentov ho využíva na vzdelávanie, nasleduje e-mail (69,45%), zábava (62,41%), práca a podnikanie (60,44%).

Používatelia prichádzajú na internet najčastejšie, aby našli informácie potrebné pre prácu a štúdium (84,73%), stiahnuť si softvér (58,17%), získať komerčné informácie (48,57%) a nájsť informáciu o tovaroch (43,74%).

Aj keď väčšina používateľov deklaruje, že vie, aký je rozdiel medzi jednoduchým a rozšíreným vyhľadávaním (77,33%), výsledky našich prieskumov ukazujú, že rozšírené vyhľadávanie sa prakticky nepoužíva. Svedčí o tom čiastočne aj odpoveď na otázku, či používatelia formulujú svoju požiadavku pomocou jedného slova, frázy alebo podľa potreby. Až 64,8% respondentov uvádza, že požiadavku formulujú podľa potreby.

Používatelia ani nepoznajú (76,01%) ani nevyužívajú (92,06%) inteligentné nástroje na vyhľadávanie, ako napríklad Lexibot, Copernic, Deep Query Manager a pod.

Pri vyhľadávaní najviac prekáža množstvo nerelevantných výsledkov (71,3%) a pomerne odmietavý postoj je aj voči množstvu reklamných bannerov (36,69%). Veľká väčšina používateľov je spokojná s nájdenými informáciami, až 51,37% používateľov udáva mieru spokojnosti 80% a 39,68% mieru spokojnosti 60%. Nespokojných je spolu iba okolo 5% používateľov. Opäť sa potvrdzuje hypotéza, že používatelia si často zamieňajú kvantitu za kvalitu. Väčšina používateľov si pozrie 2 až 5 stránok, aby našli požadovanú informáciu (56,56%) a až 40,63% používateľov je na internete denne viac ako jednu hodinu.

Výsledky prieskumu ukazujú, že je potrebné venovať väčšiu pozornosť informačnej architektúre vyhľadávacích nástrojov a pri návrhu interfejsu by sa malo vychádzať z analýzy správania sa používateľov.

Pretože internet je neriadený a index vyhľadávacích termínov je obrovský, je vhodné pomôcť používateľovi pri formulácii rešeršnej požiadavky nápovedou podobných termínov, pojmov a tém. AltaVista to napríklad rieši nápovedou **Refine your Search with AltaVista Prisma**.

---

altavista™ Web Image MP3/Audio Video Directory News Advanced Family Filter: off Settings

"internet strategy" FIND More Precision

SEARCH:  Worldwide or  Select a country RESULTS IN:  All languages  English

Refine your search with AltaVista Prisma Click a term to focus your search. Click >> to replace your search. Help

<a href="#">Internet Commerce &gt;&gt;</a>	<a href="#">Internet Strategy Consulting &gt;&gt;</a>	<a href="#">Competition &gt;&gt;</a>	<a href="#">Prospects &gt;&gt;</a>
<a href="#">Internet Presence &gt;&gt;</a>	<a href="#">Marketing Strategy &gt;&gt;</a>	<a href="#">Creation &gt;&gt;</a>	<a href="#">Publisher &gt;&gt;</a>
<a href="#">Internet Services &gt;&gt;</a>	<a href="#">Web Strategy &gt;&gt;</a>	<a href="#">John &gt;&gt;</a>	<a href="#">Strategies &gt;&gt;</a>

---

Obr. 2 Možnosť reformulácie rešeršnej požiadavky v AltaVista

K vynikajúcej službe v globálnom meradle patrí služba Wordtracker (<http://www.wordtracker.com>), ktorá pomáha pri výbere selekčných termínov na základe analýzy rešeršných požiadaviek v internete. Problém je v tom, že pri obrovskom množstve informácií v internete dostaneme pri príliš všeobecnej požiadavke aj niekoľko stotisíc záznamov. Na nasledujúcom obrázku vidíme pomoc systému pri hľadaní príbuzných termínov k pojmu "search strategy". Systém mi ponúkol 15 príbuzných termínov. Kliknutím na termín **search strategy** som získala prehľad o 15 najčastejšie rešeršných požiadavkách, v ktorých sa vyskytoval uvedený pojem. Pri analýze sa vychádzalo z 322 miliónov rešeršných požiadaviek za posledných 60 dní.

**Thesaurus**

Click on any single keyword below to query our database of over 30 million searches. Results will appear in the right hand window.

1. [rebuying](#)
2. [purchasing power](#)
3. [purchasing](#)
4. [reciprocal trade](#)
5. [repurchase](#)
6. [retailing](#)
7. [retail](#)
8. [restraint of trade](#)
9. [purchase](#)
10. [promotion](#)
11. [mercantile business](#)

Keyword	Count	Predict	Dig
website promotion	498	483	✎
promotion	444	430	✎
promotions	382	370	✎
web site promotion	365	354	✎
proforma promotions	313	303	✎
promotion codes	311	301	✎
free website promotion	310	300	✎
health promotion	229	222	✎
site promotion	209	203	✎
search engine promotion	195	189	✎
buy.com promotion code	184	178	✎
american petroleum promotions	171	166	✎
web promotion	148	143	✎

Keywords in basket: 0 / 30  
Clear All

Obr. 3 Analýza príbuzných pojmov a rešeršných požiadaviek na termín search strategy

Je ideálne, a výskum smeruje k tomu, aby sme rešeršnú požiadavku mohli formulovať v prirodzenom jazyku. Už existujú vyhľadávacie nástroje, ktoré túto funkciu podporujú. K najznámejším a aj často využívaným patrí Ask Jeeves (<http://www.askjeeves.com>).

### Výber vhodného vyhľadávacieho nástroja internetu

Na kvalitu rešerše má veľký vplyv výber vhodného vyhľadávacieho nástroja. Presvedčili sme sa o tom aj pri experimente, kde sme zadali tú istú požiadavku viacerým vyhľadávacím nástrojom.



Prakticky neexistujú dva totožné vyhľadávacie nástroje. Iné pokrytie majú adresáre, iné prieskumové stroje, v ostatnom čase sa stále viac hovorí o neviditeľnom webe, ktoré obsahuje veľa cenných informácií. Záleží od erudovanosti, ale aj skúsenosti informačného špecialistu, pre ktorý z nástrojov sa rozhodne. Všeobecne platí, že treba vyskúšať viac vyhľadávacích nástrojov.

Odporúča sa využiť aj metaprieskumové stroje, ktoré umožňujú vyhľadávanie v databázach viacerých prieskumových strojov súčasne a zároveň kombinujú výsledky vyhľadávania pri odstránení duplicitných záznamov. V súčasnosti je k dispozícii veľké množstvo metaprieskumových strojov s rôznymi funkčnými možnosťami. Odporúčam využiť **Vivissimo**, ktorý získal za rok 2002 prestížne ocenenia časopisu Search Engine Watch a to **Best Meta Search Engine**. Vivissimo (<http://www.vivissimo.com>) vychádza z výsledkov interdisciplinárneho výskumu katedry počítačovej vedy na Carnegie Mellon University v Pittsburgu. Jadrom technológie Vivissimo je zhlukovanie dokumentov, automatická organizácia dokumentov do účelných skupín.

Technológia **Vivissimo Clustering Engine™** môže využiť na vyhľadávanie rôzne databázy, výsledky organizuje do príbuzných kategórií. Výhody uvedenej technológie sú predovšetkým v používateľskom komforte a efektívnosti. Pre každú požiadavku sa vytvoria príbuzné zhľuky dokumentov, ktoré si používateľ prezrie podľa záujmu. Tým je väčšia šanca aj pre menej relevantné dokumenty, ktoré môžu mať v rámci určitého zhľuku vysoké umiestnenie.

## Kombinácia selekčných termínov pri vyhľadávaní

Výsledky pravidelných prieskumov ukazujú, že používatelia pri prieskume nevedia a nekombinujú jednotlivé termíny. Pomerne zložité používanie Boolovských operátorov ovládajú informační profesionáli, ktorí vyhľadávajú v komerčných databázových centrách. Pre bežných používateľov internetu sú ideálne preddefinované formuláre umožňujúce jednoduchým spôsobom spájať výrazy pomocou operátorov Boolovej algebry.

Výsledky prieskumu ukazujú, že iba veľmi málo používateľov využíva túto možnosť. Na otázku **Ako vyhľadáвате?** pri prieskume uskutočnenom spoločnosťou NPD New Media Services v roku 2000 odpovedalo iba 17,9% používateľov, že využíva preddefinované formuláre. 44,8 % používateľov formuluje svoju požiadavku ako frázu a ešte stále 28,6% používateľov použije iba jedno slovo. Prieskumu sa zúčastnilo 33 000 respondentov (Sullivan 2000).

Priame využitie Boolovských operátorov umožňuje vytvárať ďaleko zložitejšie požiadavky ako preddefinované formuláre. AltaVista dáva používateľovi možnosť voľby, ktorý spôsob použije. Skúsenejší používatelia môžu do formulára napísať aj takúto rešeršnú požiadavku:

**((title:peanut NEAR butter) AND (jelly OR jam) AND NOT recipe)**

Pri kombinácii selekčných termínov treba vychádzať z používateľskej podpory vyhľadávacieho nástroja. Aj keď sa niektorí odborníci snažia presadiť unifikovaný prístup k formulovaniu rešeršnej požiadavky, ako napr. **Search Engines Standardization Project**, prax je bohužiaľ zatiaľ iná. Problém je v rozdielnych funkčných možnostiach vyhľadávacích nástrojov a v rozdielnej syntaxi príkazov. Zatiaľ čo spomínaný Search Engines Standardization Project rieši nekompatibilitu v syntaxi príkazov, rozdielne funkčné možnosti sú výsledkom súperenia vyhľadávacích nástrojov o najväčší počet používateľov.

## Kritériá ovplyvňujúce samotný prieskum

Cieľom každého informačného prieskumu je na základe naformulovanej požiadavky nájsť informácie relevantné k tejto požiadavke. Z toho vyplýva, že výsledky prieskumu závisia od kvality

a spôsobu indexovania databázy, jej veľkosti, mierou zhody medzi selekčným obrazom záznamu a rešeršnou požiadavkou a pod.

## 1. Veľkosť a aktualizácia databázy

Aj keď mnohí odborníci zastávajú názor, že najväčší index ešte neznamená najlepšie výsledky vo vyhľadávaní, od roku 1995 jednotlivé prieskumové stroje súťažia o čo najväčší index. Podľa údajov z apríla 2003 najväčší index má **Google**, ktorý indexoval vyše 3 miliárd webových stránok. Na kvalitu indexu má veľký vplyv, ako často a do akej hĺbky je robot naprogramovaný, aby indexoval internet. V súčasnosti je ale ďaleko viac informácií v hĺbkovom webe alebo neviditeľnom webe. Chris Sherman a Gary Price v publikácii *The Invisible Web* (2001, s. 57) definujú neviditeľný web ako textové stránky, súbory, alebo ďalšie informácie prístupné prostredníctvom WWW, ktoré prieskumové stroje nedokážu (vzhľadom na technické obmedzenia) alebo nechcú zahrnúť do svojho indexu. Často sa používajú aj synonymné termíny hĺbkový web, alebo tmavý internet. V auguste 2001 obsahoval neviditeľný web viac ako 550 miliárd dokumentov. Viac informácií o hĺbkovom webe je v publikácii *Soni Makulovej* (2002).

Veľkosť databázy závisí aj od stratégie strojov na akceptovanie nových stránok. Stále viac prieskumových strojov kladie dôraz na kvalitu záznamov, ktorá sa často odvodzuje od analýzy kvality spojení vedúcich na stránku. Väčšina vyhľadávacích nástrojov prechádza na platené umiestnenie pre stránky komerčného charakteru. Zdôvodňujú to vyššou kvalitou služieb spočívajúcou v rýchlejšom dodaní stránky do indexu a v častejšej aktualizácii indexu.

## 2. Informačná architektúra

Internet je predovšetkým **publikačné médium**. Ide o to, akým spôsobom sa obsah dostane k čitateľovi. V súvislosti s online publikovaním na internete vzniká nová disciplína **informačná architektúra**, ktorá sa zaoberá mnohými problémami súvisiacimi s organizáciou a usporiadaním obsahu v internete. Zložitosť informačnej architektúry závisí od množstva obsahu. Inú informačnú architektúru potrebuje webové sídlo s 20 dokumentami a inú webové sídlo s niekoľko tisíc dokumentami. Cieľom informačnej architektúry je organizovať obsah do súdržnej a logickej štruktúry tak, aby sa dal jednoducho manažovať organizáciou, ktorá obsah publikuje a aby ho mohol využiť jednoduchým spôsobom používateľ, ktorý ho potrebuje.

## 3. Zmena algoritmu na nájdenie relevantných dokumentov

Počas niekoľkoročnej existencie vyhľadávacích nástrojov sa menil algoritmus výpočtu relevancie a spôsob prezentácie nájdených záznamov. Pri pomerne malom webe s relatívne statickou štruktúrou HTML dokumentov sa väčšinou brala do úvahy frekvencia počtu výskytu hľadaných termínov. Postupne so zdokonaľovaním nástrojov a pribúdaním informácií v internete sa menili aj algoritmy. Každý robot vyhľadávacieho nástroja má svoj vlastný algoritmus, ktorý sa z času na čas mení. Väčšina súčasných vyhľadávacích nástrojov využíva pri určení relevancie analýzu spojení.

## 4. Analýza spojení prieskumovými strojmi

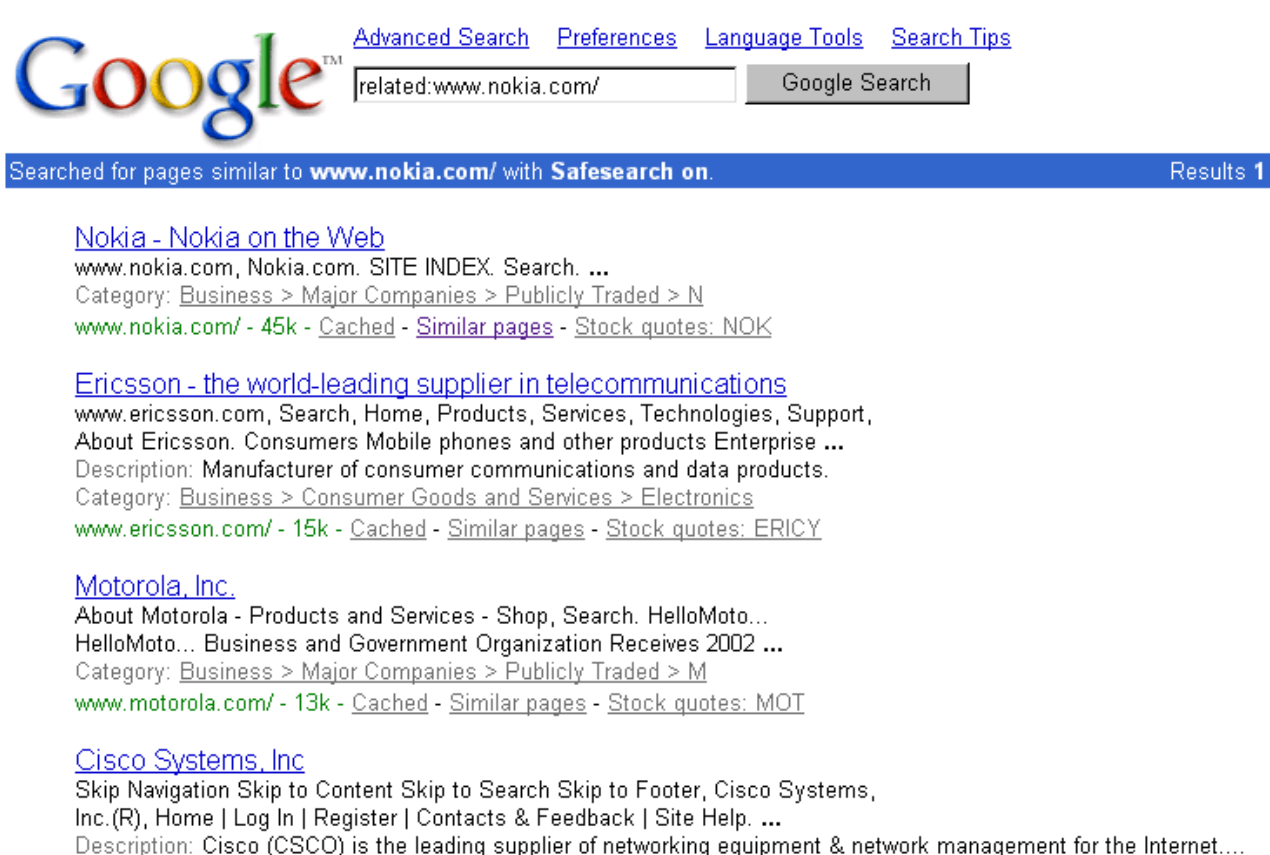
Prvé prieskumové stroje brali pri radení dokumentov výsledku rešerše tzv. **popularitu spojení**, ktoré viedli na stránku. Prakticky to znamenalo, že čím na stránku viedlo viac spojení, tým sa stránka pokladala za dôležitejšiu. Niektoré stroje indexovali iba stránky na ktoré viedlo veľa spojení, prípadne dávali týmto stránkam väčšiu váhu. Dnes sa už vie, že veľa spojení, ktoré vedú na stránku sa dá vytvoriť aj umelým spôsobom a preto sa skôr berie do úvahy aj **analýza kvality spojení**, ktoré vedú na stránku.

Pri analýze spojení sa určí dôležitosť jednotlivých spojení a tiež kontext, textová súvislosť spojení. Aj keď sa mnohí snažia oklamať prieskumové stroje, treba si uvedomiť, že najdôležitejší je obsah. Čím má stránka kvalitnejší obsah, tým viac spojení na ňu vedie a získava väčšiu váhu pri hodnotení. Dôležitý faktor je kvalita stránky, z ktorej vedú spojenia. Priekopníkom v tejto oblasti je vyhľadávací nástroj **Google** a jeho patentovaná technológia PageRank™. Ako dôležité kritériá sú analýza textu spojení, kontextu, obsahu stránok, ktoré odkazujú na túto stránku a pod.

## Zjemňovanie a ladenie rešeršnej požiadavky

Podľa množstva záznamov môžeme ďalej zjemňovať a ladiť rešeršnú požiadavku. Mnohé vyhľadávacie nástroje na to využívajú hypertextové spojenie **Podobné stránky** (v anglickej verzii Similar pages). Uvedená funkcia je mimoriadne užitočná na nájdenie stránok s podobným obsahom, získanie informácií o konkurencii a pod. Na nasledujúcom obrázku vidíte ukážku výstupnej zostavy po kliknutí na spojenie "Podobné stránky" spoločnosti Nokia (príkaz related:www.nokia.com/) v Google.

---



The screenshot shows the Google search interface. The search bar contains the query "related:www.nokia.com/". Below the search bar, there are links for "Advanced Search", "Preferences", "Language Tools", and "Search Tips". The search results are displayed on a blue background. The first result is "Nokia - Nokia on the Web" with a description of the website and a link to "Similar pages". The second result is "Ericsson - the world-leading supplier in telecommunications" with a description and a link to "Similar pages". The third result is "Motorola, Inc." with a description and a link to "Similar pages". The fourth result is "Cisco Systems, Inc." with a description and a link to "Similar pages".

### Obr. 4 Využitie zjemňovania rešerše kliknutím na spojenie "Podobné stránky"

Ako vidíme na obrázku, systém ponúkol stránky konkurencie, ale s podobným obsahom (Ericsson, Motorola, Cisco a pod.).

Nové vyhľadávacie nástroje tretej generácia, ako napríklad AskJeeves, Northern Light, Applied Semantics a SurWax vychádzajú pri vyhľadávaní z vybudovanej poznatkovej bázy, ktorú vytvorili experti. Podporuje sa využívanie tezaurov.

K ďalším trendom pri radení dokumentov na výstupe je zohľadňovať ľudský faktor, ktorý vychádza zo správania a posudzovania webových stránok miliónmi používateľov. Tento prístup analýzy

spojení na stránkach využívajú hlavne Google, Direct Hit a Teoma. Teoma pri analýze spojení identifikuje tzv. "komunity" na webe, a potom určí, ktoré zo sídiel sú authority pre tieto community. Podobný prístup má aj Google, ale zatiaľ čo Google na určenie relevantných dokumentov využíva tzv. kolektívnu múdrosť celého webu, Teome pomôžu zistiť najlepšie stránky authority jednotlivých komunit

Zjemňovanie rešeršnej požiadavky v množine nájdených záznamov patrí k užitočným funkciám, ktoré poskytujú iba niektoré nástroje. Na nasledujúcom obrázku vidíte, ako to rieši Google.



Dissatisfied with your search results? [Help us improve.](#)

Obr. 5 Požiadavku môžeme ďalej modifikovať postupným pridávaním pojmov.

K veľmi účinným nástrojom na zjemňovanie a ladenie rešerše patria preddefinované formuláre umožňujúce nastaviť celý rad filtrov. Preddefinované formuláre ponúka väčšina zahraničných vyhľadávacích nástrojov.

Väčšina prvých prieskumových strojov poskytovala výsledok rešerše vo forme dlhého vertikálneho zoznamu. Súčasný vyhľadávacie nástroje umožňujú ďalej pracovať s výsledkami rešerše, prezrieť si systém pojmov, kategórií a ďalej pracovať s týmito kategóriami. Príkladmi môžu byť [Northern Light](#), [Query Server](#) a [Vivisimo](#).

Väčšina prieskumových strojov uzatvára strategické spojenia s adresármi, čím sa dopĺňa ľudský aspekt a expertíza pri vytváraní adresárov so sofistikovanými možnosťami prieskumových strojov. Výsledky rešerše z indexu prieskumových strojov sú dopĺňané záznamami s adresárov.

Horeuvedené príklady dokumentujú, že súčasné vyhľadávacie nástroje tretej generácie už umožňujú zasahovať a modifikovať rešeršnú stratégiu a taktiku vo viacerých fázach prieskumového procesu tak, ako to bolo možné pri vyhľadávaní v komerčných databázových centrách. Tí, ktorí si pamätajú prvé vyhľadávacie nástroje a princíp ich práce mi určite dajú za pravdu, že sa urobil veľký pokrok. Ale splnenie vízie sémantického webu vyžaduje ešte seriózny interdisciplinárny vedecký výskum pri súčinnosti všetkých prvkov prieskumového procesu.

## Záver

Výsledky prieskumu využívania informačných zdrojov a miery používateľskej spokojnosti s vyhľadávacími nástrojmi ukazujú, že internet už pevne zapustil svoje korene aj na Slovensku a stáva sa bežnou súčasťou denného života a dôležitým zdrojom informácií. Je evidentné, že používatelia neovládajú metódy a techniky prístupu k informáciám, často vyhľadávajú intuitívne a k spokojnosti im stačí veľké množstvo záznamov. Nájdené záznamy nedokážu hodnotiť z hľadiska presnosti a úplnosti.

Je potrebné venovať väčšiu pozornosť informačnej architektúre vyhľadávacích nástrojov a pri návrhu používateľského interfejsu vychádzať z analýzy správania sa rôznych kategórií používateľov pri vyhľadávaní informácií.

V súčasnosti môžeme hovoriť o vyhľadávacích nástrojoch tretej generácie. Je snaha pomôcť používateľovi v každej fáze prieskumového procesu. Napriek tomu treba ešte veľa práce, na to, aby sa uskutočnila v praxi myšlienka sémantického webu tak, ako ju formuloval Tim Berners Lee. „Sémantický web je rozšírením dnešného webu, v ktorom bude informáciám daný jasne definovaný význam, lepšie umožňujúci kooperáciu počítačov a ľudí.“ (Berners-Lee, 2001). Tieto problémy sú spôsobené skutočnosťou, že dáta a informácie sú na dnešnom webe uložené v strojom čitateľnej, no nie strojom zrozumiteľnej forme, čo spôsobuje neschopnosť strojov zachytiť sémantiku uvedených dát.

Veľké rezervy sú aj v systéme vzdelávania používateľov týkajúce sa vyhľadávania informácií v digitálnom priestore. Treba si uvedomiť, že množstvo informácií v digitálnej forme narastá zároveň aj s informačným preťažením používateľov. Netreba zabúdať na to, že informáciu, ktorú nikto nenájde ani nikto neprečíta.

## Literatúra

Agichtein, Eugene, Lawrence, Steve, Gravano, Luis. 2000. Learning Search Engine Specific Query Transformations for Question Answering. In WWW10, May 2-5, 2001, Hong Kong. [cit. 2003-03-13]. Dostupné na internete: <<http://www10.org/cdrom/papers/348/>>.

Berners, Lee a kol. 2001. The Semantic Web. In Scientific American, May 2001, . [cit. 2003-03-13]. Dostupné na internete: <<http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>>.

Catledge, L. D, Pitkow, J. E. 1995. Characterizing Browsing Strategies in the World Wide Web. [cit. 2003-03-13]. Dostupné na internete: <<http://www.igd.fhg.de/www/www95/papers/80/userpatterns/UserPatterns.Paper4.formatted.html>>

DELOS brainstorming report. 2001. San Cassiano, Alta Badia, Italy, June 2001, ERCIM-02-W02. [cit. 2003-03-13]. Dostupné na internete: <<http://delos-noe.iei.pi.cnr.it/activities/researchforum/Brainstorming/brainstorming-report.pdf>>.

Engel, Marc. 2001. “Sex” Popular on the Web, Many People Inefficient at Reaching Their Online Destinations. In: Alexa Research News Release. February 14, 2001.

Fuller, Rodney, De Graff, Johannes. Measuring User Motivation from Server Log Files. [cit. 2003-03-13]. Dostupné na internete: <<http://www.microsoft.com/usability/webconf/fuller/fuller.htm>>.

Griffiths, R., Brophy, P. 2002. Student searching behaviour in the JISC Information Environment. In Ariadne Issue 33. [cit. 2003-03-13]. Dostupné na internete: <<http://www.ariadne.ac.uk/issue33/edner/intro.html>>.

Hock, Randolph. 2001. The Extreme Searcher's Guide to Web Search Engines: A Handbook for the Serious Searcher. Medford : Information Today, 2001. 241s.

Huberman, Bernardo, A., Priolli, Peter, T., Pitkow, James, E., Lukose, Rajan. M. 1998. Strong Regularities in World Wide Web Surfing. In *Science*, vol. 280, č. 5360, s. 95-97. [cit. 2003-03-13]. Dostupné na internete: <<http://citeseer.nj.nec.com/cache/papers/cs/18888/http://zSzzSzexternal.nj.nec.com/zS~gileszSzhubermanzSz98.huberman.pdf/huberman98strong.pdf>>.

Jansen, Bernard. J. 2000. The effect of query complexity on Web searching results. In Information Research [online]. 2000, vol. 6, no.1 [cit. 2003-02-20]. Dostupné na internete: <<http://www.shef.ac.uk/~is/publications/infres/paper87.html>>

- Jansen, B., Spink, A., Saracevic, T. 2000. Real life, real users, and real needs: A study and analysis of user queries on the web. In *Information Processing and Management*, 2000, vol. 36, no. 2, s. 207-227.
- Jansen, B. J., Spink, A., Bateman, J., & Saracevic, T. 1998. Real life information retrieval: A study of user queries on the web. In *SIGIR Forum*, 1998, vol. 32. no. 1, pp. 5 -17. [cit. 2003-02-20]. Dostupné na internete: <<http://jimjansen.tripod.com/academic/pubs/sigirforum98/forum98.html>>
- Lyman, Peter - Varian, Hal R. 2000. *How Much Information?* University of California 2000. [cit. 2003-03-13]. Dostupné na internete: < <http://www.sims.berkeley.edu/how-much-info>>.
- McGovern, Gerry - Norton, Rob. 2002. *Content Critical*. London : Pearson Education Limited, 2002. 241 s.
- Makulová, Soňa. 2002. *Vyhľadavanie informácií v internete. Problémy, východiská, postupy*. Bratislava : EL&T, 2002. 376 s. ISBN 80-88812-16-X.
- Makulová, Soňa. 2003. *Informačné správanie používateľov pri vyhľadávanií informácií v internete*. In: *Zborník z medzinárodnej konferencie konanej v rámci programu DELOS - CEE. Informačné správanie a digitálne knižnice*. v tlači.
- Rosenfeld, Louis - Morville, Peter. 2002 *Information Architecture for the World Wide Web*. Sebastopol : O'Reilly & Associates, Inc. 2002. 461 s.
- Search Engines Continue to be an Important Method For Discovering New Sites and Content. 2003. In *StatMarket* 12. march 2003. [cit. 2003-03-13]. Dostupné na internete: <[http://www.statmarket.com/cgi-bin/sm.cgi?sm&feature&week\\_stat](http://www.statmarket.com/cgi-bin/sm.cgi?sm&feature&week_stat)>.
- Sherman, Chris. 2002. 2001's Most Wanted Search Terms. In: *SearchDay*, January 2, 2002 - Number 172-73. [cit. 2003-03-13]. Dostupné na internete: <<http://searchenginewatch.com/searchday/02/sd0102-words.html>>
- Sherman, Chris - Price, Gary. 2001. *The Invisible Web*. Medford : Information Today, Inc., 2001. 439 s.
- Siegfried, S., Bates, M., Wilde, D. 1993. A profile of end-user searching behavior by humanities scholars: the Getty online searching project report no. 2. In *Journal of the American Society for Information Science*, 1993, vol. 44, no.5, s. 273-291.
- Spink, A. a kol. 1998. Modeling users successive searches in digital environments. *D-Lib Magazine*, April 98. [cit. 2003-03-13]. Dostupné na internete: < <http://www.dlib.org/dlib/april98/04spink.html>>
- Spink, A. a kol. 1998. Modeling users successive searches in digital environments. *D-Lib Magazine*, April 98. [cit. 2003-03-13]. Dostupné na internete: < <http://www.dlib.org/dlib/april98/04spink.html>>
- Sullivan, Danny. 2003. Searches Per Day. In *SearchEngineWatch*. [cit. 2003-03-13]. Dostupné na internete: <http://searchenginewatch.com/reports/perday.html>
- Sullivan, Danny. 2002. Jupiter Media Metrix Search Engine Ratings. [cit. 2002-03-20]. In: *Search Engine Watch*, 4 March, 2002. URL: <http://searchenginewatch.com/subscribers/submit-dir.html>.
- Tauscher, Linda - Greenberg, Saul. 1997. How People Revisit Web Pages: Empirical Findings and Implications for the Design of History Systems. In *International Journal of Human-Computer Studies*, 1997, vol. 47, s. 97-137. [cit. 2003-03-13]. Dostupné na internete: <<http://ijhcs.open.ac.uk/tauscher/tauscher.pdf>>.

The UCLA Internet report. Surveying the Digital Future. Year Three. 2003. UCLA Center for Communication Policy. February 2003. 87 s.

Publikácia bola spracovaná v rámci vedeckého grantového projektu VEGA 1/9236/02 Interakcia človeka s informačným prostredím..