

Návrh riešenia problémov pri vyhľadávaní informácií v internete alebo Od kvantity ku kvalite

Soňa Makulová

Filozofická fakulta Univerzita Komenského, Bratislava
makulova@eiet.sk



Prof. PhDr. Soňa Makulová, Ph.D., prednáša na katedre knihovníckej a informačnej vedy Filozofickej fakulty Univerzity Komenského v Bratislave. Zabýva sa teoreticko-metodologickými aspektmi zavádzenia moderných informačných technológií v knihovnícko-informačných systémoch, informačným prúzkumom v prostredí globálnych informačných sietí, externými informačnými zdrojmi a informačnou architektúrou. Je autorkou kníh *Automatizácia knižníc*, *Spríevodca po Internete* a *Internet v riadení a obchode firmy* (spoluautor Vladimír Burčík), *Spríevodca po Internete alebo Internet od A po Z*, *Internet do vrečka* a *Vyhľadávanie informácií v internete*. Napsala rovněž skripta *Bázy dát v sústave VTEI* a *Úvod do informačných technológií II* (spoluautor Štefan Kimlička) a doma a v zahraničí publikovala přes 140 štúdií, recenzií a článkov. Je prezidentkou národnej multimediálnej asociácie *SlovakPrix Multimedia* a od roku 2003 je členom svetovej poroty *Grand Jury* pro hodnotenie multimédií. Prednášala na zahraničných konferenciách a univerzitách v Českej republike, Dánsku, Finsku, Rakúsku, Hongkongu, Maďarsku, Nemecku, Poľsku, Rakúsku, Španielsku, Švédsku, USA a Veľkej Británii.

1. Úvod

Súčasný prienik komunikačných a informačných technológií do všetkých oblastí nášho života spôsobuje obrovský nárast informácií v digitálnej forme. Tu si treba uvedomiť, že podobne ako nedostatok aj príliš veľa informácií môže škodiť. Presvedčili nás o tom aj udalosti z 11. septembra, keď vláda USA nevedela útoky predpovedať nie z nedostatku informácií, ale pretože ich bolo príliš veľa a nedokázala ich správne vyhodnotiť a interpretovať.

Aj keď sa v internete nachádza veľa cenných a hodnotných informácií, prieskumy využívania internetu ukazujú, že väčšina používateľov považuje za jeden z najväznejších problémov súčasného internetu nájsť kvalitnú a relevantnú informáciu. Cieľom štúdie je na základe analýzy problémov, pred ktorými stoja v súčasnosti používatelia informácií a projektanti vyhľadávacích strojov, navrhnúť riešenie uvedených problémov.

2. Význam internetu pri budovaní informačnej spoločnosti

Na nedávnom celosvetovom sumite o informačnej spoločnosti, ktorý sa konal v decembri v roku 2003 v Ženeve, sa deklarovala naša spoločná túžba a záväzok vybudovať informačnú spoločnosť, kde môže každý tvoriť, získavať, používať a deliť sa o informácie a poznatky. V deklarácii sa zdôraznil význam informačných a komunikačných technológií, predovšetkým internetu, ktorý zohral kľúčovú úlohu pri budovaní informačnej spoločnosti v celosvetovom meradle.

Internet by sme si mohli stručne charakterizovať ako informačné, komunikačné, obchodné, marketingové a reklamné médium, ale aj ako globálnu sieť, prostredníctvom ktorej je dnes

prepojených okolo 285 miliónov hostiteľských počítačov s viac ako 934 miliónmi používateľov na celom svete. Internet úplne zmenil model komunikácie medzi ľuďmi, a preto jeho význam mnohí srovnávajú s významom vynálezu knihtače Jánom Gutenbergom.

Analytická sekcia britského týždenníka The Economist v spolupráci s IBM Institute for Business Values uverejnila štúdiu The 2004 e-readiness rankings (2004), ktorá hodnotí krajiny z hľadiska úrovne využívania internetu vo všetkých oblastiach života. Na prvých miestach sa opäť objavili všetky škandinávské krajiny, pričom prvým v celkovom poradí je Dánsko, ktoré získalo z možných 10 bodov 8,28. V prvej desiatke je ešte Veľká Británia (2), USA (klesly z 3. miesta na 6. miesto), Singapore (7), Holandsko (8), Hongkong (9) a Švajčiarsko (10). Pre zaujímavosť z nových členských krajín EU je na tom najlepšie Estónsko (26. miesto), hneď za ním je Česká republika (27), Maďarsko (30), Slovinsko (31), Litva (34), Poľsko (36), Lotyšsko (38) a na poslednom mieste je Slovenská republika (39).

Autori štúdie konštatujú, že kľúčom k úspechu je koordinácia a spoločné programy vládnych inštitúcií a IT priemyslu. V tejto oblasti sú vynikajúce predovšetkým škandinávské krajiny, kde sme svedkami prudkého nárastu penetrácie rýchleho pripojenia k internetu a s tým spojených nových služieb, ako aj tlaku na odbúranie byrokratického styku s úradmi a budovanie e-governmentu.

Pri budovaní informačnej spoločnosti má v zmysle základných materiálov EU význam predovšetkým rýchly internet. Tak môžu používatelia využívať video, vzdelávacie služby, turistické, geografické informácie apod.

Aj keď podľa posledného výskumu Internet v Slovenskej republike – 2/2004 spoločnosti Taylor Nelson Sofres využívalo internet v júli 2004 takmer 29,2% obyvateľov (v porovnaní s 5% v roku 1997 podľa Net Projektu), k najväčším bariéram patrí ešte pomerne vysoká cena pripojenia k internetu a počítačov. Podľa vyhodnotenia programu eEurope+ na Slovensku stojí pripojenie k internetu 15% priemerného mesačného príjmu, pre porovnanie v Čechách je to 5,5% a v Slovinsku dokonca iba 2,2%. Podobne je to aj s cenou počítačov, zatiaľ čo na Slovensku stojí priemerný počítač 164% priemerného príjmu, v Čechách stačí iba 146% a v Slovinsku iba 86% (Topol 2004). Ďaleko lepšia situácia v Čechách je výsledkom koordinovaného prístupu Ministerstva informatiky, zatiaľ čo na Slovensku sa problematika informatizácie presunula z Ministerstva školstva na Ministerstvo dopravy, pôšt a telekomunikácií a aj napriek prísľubom vlády stále chýba systematický a koordinovaný postup pri budovaní jednotlivých pilierov informačnej spoločnosti.

3. Problémy pri vyhľadávaní informácií v internete

Pri analýze problémov vyhľadávaní informácií v internete sa sústreďujeme na dva hlavné okruhy problémov. Na jednej strane sú to problémy, ktoré musia riešiť samotní používatelia informácií a na druhej strane problémy, ktoré musia riešiť projektanti a tvorcovia vyhľadávacích nástrojov.

3.1 Problémy používateľov informácií

Problémy používateľov informácií by sme mohli zhrnúť nasledovne:

- informačný prieskum na webe je iný ako v databázových centrách a digitálnych knižniciach,
- množstvo vyhľadávacích nástrojov,
- informácie sú rozptýlené v hĺbkovom a povrchovom webe,
- používatelia nevedia správne formulovať svoju informačnú požiadavku a hodnotiť nájdené záznamy.

3.1.1 Formulovanie informačnej požiadavky

Ak si položíme otázku, prečo nie sú používatelia internetu spokojní s výsledkom nájdených záznamov v internete, nájdeme viac objektívnych a subjektívnych príčin. K subjektívnym príčinám patrí predovšetkým nesprávna formulácia rešeršnej požiadavky. Veľa používateľov nepozná možnosti vyhľadávacích nástrojov a ani nemá dostatočné vedomosti na to, aby správne dokázali používať výroky Boolovej algebry. Iba málo vyhľadávacích nástrojov používa preddefinované formuláre umožňujúce aj neskúseným používateľom naformulovať zložitú rešerš s rôznymi filtermi. O tom, že informačný prieskum na webe je iný ako v databázových centrách a digitálnych knižniciach, svedčia aj výsledky štúdie Jansena a Poocha (2000).

V nasledujúcej tabuľke vidíme porovnanie formulovania informačných požiadaviek v digitálnom priestore.

Ako formulujú používatelia svoju informačnú požiadavku pri vyhľadávaní informácií v digitálnom priestore

	Web	Online DC	OPAC
Priemerný počet požiadaviek	1–2	7–16	2–5
Počet termínov v požiadavke	2	6–9	1–2
Počet prezretých dokumentov	<=10	Približne 10	<50
Požiadavky využívajúce zdokonalené vyhľadávanie	9%	9%	8%
Požiadavky s Boolovskými operátormi	8%	37%	1%

Ako vidíme, formulovanie informačnej požiadavky je v digitálnom priestore odlišné. Vyhľadávanie v databázových centrách využívajú väčšinou informační profesionáli, experti vo vyhľadávaní, čomu zodpovedá aj vyšší počet požiadaviek s Boolovskými operátormi, väčší počet termínov v požiadavke a vyšší priemerný počet požiadaviek na jednu reláciu. Požiadavky s Boolovskými operátormi sa najmenej využívajú pri vyhľadávaní v on-line katalógoch knižníc a najviac v databázových centrách.

Výsledky štúdií správania používateľov treba zohľadniť pri návrhu interfejsu pre vyhľadávanie tak, aby sa zohľadnilo správanie všetkých skupín používateľov (nováčikov, expertov apod.). Veľa podnetných informácií týkajúcich sa užívateľského rozhrania a problematiky vyhľadávania informácií je v prácach Richarda Papíka (2001).

3.1.2 Výber vyhľadávacieho nástroja

K ďalším subjektívnym príčinám patrí nesprávny výber vyhľadávacieho nástroja. Situáciu komplikuje množstvo rôznych vyhľadávacích nástrojov a skutočnosť, že neexistuje nástroj, ktorý by komplexne pokrýval celý web. V súčasnosti sa pri vyhľadávaní informácií v internete využívajú predovšetkým nasledovné kategórie vyhľadávacích nástrojov:

Prieskumové stroje, alebo fulltextové vyhľadávače, ktoré využívajú na tvorbu databázy program tzv. robot (spider, webcrawler apod.). K typickým predstaviteľom patria napríklad

Google, Teoma, AltaVista apod. Pre zaujímavosť vyhľadávač Google indexoval v januári 2005 viac ako 8 miliárd webových stránok.

Predmetové adresáre pokrývajú ďaleko menšiu časť webovského priestoru ako prieskumové stroje, pretože sa databáza nevytvára automaticky, ale vytvárajú ju redaktori. Ich výhodou je prísna hierarchická klasifikácia a celý rad ďalších kategórií a podkategórií. Používatelia, ktorí využívajú tieto systémy, by mali vedieť, čo hľadajú v internete a do ktorej predmetovej kategórie patrí hľadaná informácia. K najznámejším predmetovým adresárom patria Yahoo!, Open Directory Project na Slovensku a v Čechách je to Seznam. V januári 2005 obsahoval tzv. ľudský adresár webu Open Directory Project vyše 4 miliónov sídiel vo viac ako 590 000 kategóriách s využitím práce 66 270 externých redaktorov. Čiže databáza predmetového adresára ODP bola takmer 2000krát menšia ako databáza Google.

Metaprieskumové stroje (metasearch engines) alebo paralelné vyhľadávacie nástroje predstavujú nadstavbu tradičných prieskumových strojov. Ich základnou úlohou je umožniť vyhľadávanie v databázach viacerých prieskumových strojov súčasne a zároveň kombinovať výsledky vyhľadávania v rôznych prieskumových strojoch pri odstránení duplicitných záznamov. Výhodou je jednotný používateľský interfejs. K najznámejším patria Vivisimo, Metacrawler, Dogpile a Copernic. Ich nevýhodou je, že vo výsledkoch vyhľadávania často nie sú zreteľne odlišené platené záznamy, čo experti vo vyhľadávaní považujú za neetické.

Virtuálne knižnice a predmetové brány predstavujú kvalitné metainformačné systémy, ktoré vytvárajú väčšinou profesionálni knihovníci špecializovaní na vyhľadanie kvalitného obsahu. Pri kategorizácii stránok sa využívajú osvedčené klasifikačné schémy organizácie poznania. K vysokokvalitným a odporúčaným patria Librarians' Index To The Internet, Infomine, AcademicInfo, BUBL LINK, About.com, Internet Public Library, Resource Discovery Network, Jednotná informačná brána apod. V týchto adresároch sa nachádzajú systematicky spracované a selektované informačné zdroje, ktoré navyše obsahujú pridanú hodnotu buď v podobe recenzie, anotácie, predmetového triedenia, krížových odkazov apod. Profesionálna tvorba je zárukou ich kvality a spoľahlivosti.

Informačné systémy o neviditeľnom webe monitorujú web, ktorého obsah prieskumové stroje nedokážu alebo z finančného hľadiska nezahŕňajú do svojho indexu. K najznámejším systémom patria The Invisible Web Directory, Complete Planet, Librarians' Index To The Internet, Infomine, AcademicInfo apod.

Univerzálne, respektíve globálne adresáre nestačia monitorovať nárast internetu, a preto sa stávajú stále viac **populárne špecializované predmetové adresáre**. V odbornej literatúre sa nazývajú aj vortály.

K dôležitým kritériám pri výbere vhodného vyhľadávacieho nástroja ďalej patria:

- veľkosť indexu vyhľadávacieho nástroja (prieskumové stroje majú ďaleko väčší index),
- spôsob indexovania webovských stránok a radenia záznamov na výstupe,
- možnosti vyhľadávania (Booleovské operátory, proximitné operátory, pravostranné rozšírenie, filtre apod.),

- aké typy dokumentov pokrýva vyhľadávací nástroj (správy, časopisy, mapy, elektronické konferencie apod.),
- aké typy formátov sú v databáze (html dokumenty, PDF, excell apod.),
- používateľská podpora a prívetivosť,
- možnosť personalizácie vyhľadávania apod.

Odporúčania pre voľbu vhodného vyhľadávacieho nástroja internetu a správnej rešeršnej stratégie sú uvedené v publikácii **Vyhľadávanie informácií v internete: problémy, východiská, postupy** (Makulová, 2002).

Záver roku 2004 bol poznamenaný rozmachom v oblasti personalizácie vyhľadávacích nástrojov. Beta verzie personalizácie predstavili Google, Ask Jeeves, Microsoft Network a Yahoo!. Sofistikovaná personalizácia predstavuje jednu z možností riešenia problémov kvality výsledkov vyhľadávania.

3.1.3 Neviditeľný web

K objektívnej príčine, o ktorej sa v poslednom čase stále viac píše v odbornej literatúre, patrí tzv. neviditeľný alebo hĺbkový web. Podstata hĺbkového webu spočíva v tom, že súčasné fulltextové vyhľadávacie nástroje veľa informácií na webe nedokážu nájsť, teda sú pre ne neviditeľné.

Pojem neviditeľný web sa objavuje v odbornej literatúre približne okolo roku 1999, keď sa prišlo na to, že prieskumové stroje neindexujú stále viac webovského priestoru. 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 štúdiu Michaela Bergmana (2001) sa uvádza, že hĺbkový web je až 500krát väčší ako tzv. „surface“ World Wide Web (povrchový web). Podľa autora hĺbkový web obsahuje stovky miliárd dokumentov vysokej kvality, ktoré sú prístupné prostredníctvom databáz. Nie zanedbateľnou skutočnosťou je fakt, že až 95 % informácií v hĺbkovom webe patria k verejne prístupným informáciám, ktoré sú prístupné bez poplatkov.

Príčin, ktoré vedú k neviditeľnému webu, je viacero a sú uvedené v nasledujúcej tabuľke.

Typy obsahu v neviditeľnom webe

Typ obsahu v neviditeľnom webe	Príčina, prečo je obsah neviditeľný
odpojené stránky	robot nemôže sledovať spojenie na stránku
stránka obsahuje hlavne obrázky, video, audio	nedostatok textu, aby robot porozumel obsahu stránky
stránka obsahuje hlavne súbory typu PDF, Postscript, flash, komprimované súbory apod.	technicky je stránka indexovateľná, ale väčšinou sa neindexuje z ekonomických dôvodov
obsah v relačných databázach	roboty nedokážu vyplniť požadované polia v interaktívnych formulároch
obsah, ktorý sa mení v reálnom čase	obrovské množstvo dát, ktoré sa neustále mení
dynamicky vytváraný obsah	personalizovaný obsah nie je zaujímavý pre väčšinu používateľov internetu

Z hľadiska typológie môžeme hovoriť o štyroch typoch neviditeľného webu: nepriehľadný web, súkromný web, špeciálny alebo vlastnícky web a skutočne neviditeľný web (Makulová, 2002).

3.1.4 Informačné potreby používateľov internetu

Prieskumy používateľov pri vyhľadávaní informácií v internete ukazujú, že používatelia okrem toho, že nepoznajú vyhľadávacie nástroje internetu, nevedia správne formulovať rešeršnú požiadavku a hodnotiť nájdené záznamy. Jeden z najkompletnejších prieskumov týkajúci sa správania používateľov uskutočnili Amanda Spink, Dietmar Wolfram, Tefko Saracevic a Bernard Jansen v rokoch 1997 až 2001. Pri prieskume vychádzali z analýzy transakčných logov na vyhľadávacích serveroch Excite, AlltheWeb a Ask Jeeves, neskôr aj z AltaVista. Cieľom výskumov bolo zistiť správanie používateľov tak, aby sa výsledky dali využiť pri zdokonaľovaní vyhľadávacích nástrojov, interfejsu a služieb.

Výsledky výskumu poukazujú na niektoré zaujímavé skutočnosti:

- väčšina požiadaviek má v priemere 2,8 termínov, v roku 1996 to bolo iba 1,5 termínov, vo všeobecnosti je ale počet termínov veľmi malý,
- jednotlivé termíny nevystihujú obsah toho, čo chce používateľ nájsť,
- používatelia nevedia používať Boolovské operátory (80 % prieskumov je bez Boolovských operátorov),
- v roku 2001 si iba tretina používateľov pozrela záznamy na ďalšej obrazovke (Spink, Jansen, Wolfram, Saracevic, 2002),
- používatelia neformulujú svoju požiadavku formou otázky, aj keď to vyhľadávacie nástroj umožňuje (v prípade Ask Jeeves je to iba 50 % požiadaviek),
- 78 % požiadaviek sa ďalej nemodifikuje, nevyužíva sa možnosť spätnej väzby,
- 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.

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 vyhľadávanie 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.

K zaujímavým zisteniam patrí aj posun vo vyhľadávaných témach. Zatiaľ čo v roku 1997 sa predovšetkým hľadali stránky týkajúce sa zábavy a sexu, v roku 2001 to boli stránky so zameraním na obchod, zamestnanie, ekonomiku, ľudí a cestovanie. Súvisí to aj s celkovou orientáciou webu smerom k obchodu (Spink, Jansen, Wolfram, Saracevic, 2002).

Problém je v tom, že 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 vyhľadávací nástroj interpretuje požiadavku, aké sú rozdiely medzi vyhľadávacími nástrojmi v internete a v databázových centrách, akým spôsobom sa používajú operátory Boolovej algebry apod. Objavujú sa aj názory, že vyhľadávacie nástroje internetu sa budú prispôsobovať správaniu priemerného internetového používateľa (Larsen, 1977).

O tom, aké boli informačné potreby používateľov internetu vo svete za rok 2003, svedčí nasledovná tabuľka.

Najviac hľadané termíny v svetových vyhľadávačoch

Google	Yahoo!	Lycos
Britney Spears	KaZaA	KaZaA
Harry Potter	Harry Potter	Britney Spears
Matrix	American Idol	Dragonball
Shakira	Britney Spears	Paris Hilton
David Beckham	50 cent	IRS
50 cent	Eminem	Kobe Bryant
Iraq	WWE	Christmas
Lord of the rings	Paris Hilton	NFL
Kobe Bryant	NASCAR	Pamela Anderson
Tour de France	Christina Aguilera	Brooke Burke

Pre zaujímavosť, v roku 2001, keď došlo k tragickým udalostiam v USA, sa v prvej desiatke umiestnili v uvedených vyhľadávačoch termíny **World Trade Center**, **Osama bin Laden**, **Taliban**, **Anthrax** a **Nostradamus**. V januári 2005 po tragických udalostiach v južnej Ázii podľa služby Wordtracker z 11. januára je na prvom mieste slovo tsunami.

V nasledujúcej tabuľke vidíme najviac vyhľadávané slová na najnavštevovanejšom slovenskom vyhľadávači Zozname slovenského internetu za rok 2004. Slová uvádzame tak, ako boli zadané do vyhľadávača, čiže v niektorých prípadoch aj bez diakritiky.

Najviac hľadané slová na Zozname slovenského internetu

Január	Február	Marec	Apríl	Máj	Jún	Júl	August
práca	práca	práca	sex	práca	pokec	last minute	pokec
sex	sex	sex	práca	pokec	wallpapers	pokec	práca
sociálna poisťovňa	autobazár	referáty	referáty	belohorcova	brigáda	práca	last minute
referáty	porno	autobazár	autobazár	sex	práca	cestovne kancelárie	mapy
autobazár	referáty	daňový úrad	porno	autobazár	zamestnanie	mapy	horoskop
porno	obchodný register	porno	pohľadnice	porno	mapy	počasie	autobazár
sms	sms	obchodný register	obchodný register	zamestnanie	autobazár	autobazár	sex
danovy urad	cestovne kancelarie	nábytok	horoskop	referáty	porno	Chorvátsko	reality
cestovne kancelarie	mp3	sms	zoznamka	zoznamka	cestovne kancelarie	csakova	nábytok
horoskop	zamestnanie	horoskop	mapy	wallpapers	sex	sex	cestovne kancelarie
reality	nábytok	zamestnanie	sms	mapy	počasie	brigáda	zoznamka
obchodny register	mapy	mp3	erotika	horoskop	brigády	reality	počasie
zamestnanie	pohľadnice	cestovne kancelarie	inzercia	sms	zoznamka	zoznamka	slovník
erotika	reality	erotika	cestovne kancelarie	nábytok	pohľadnice	porno	porno
mp3	horoskop	reality	nábytok	cestovne kancelarie	horoskop	horoskop	Bratislava

Ako vidíme, na Slovensku na prvých miestach prevládajú kľúčové slová týkajúce sa práce, sexu, ale aj zábavy a cestovania. V čase podávania daňových priznaní sú to slová ako daňový úrad, obchodný register, v letných mesiacoch v čase dovolení slová cestovné kancelárie, mapy, počasie, last minute, brigáda apod.

Pre porovnanie uvádzame najviac hľadané slová na Slovensku za rok 2001 na portáloch Atlas, Centrum a Surf.sk. Je zaujímavé, že k najviac hľadaným slovám patrilo slovo SMS. Bolo to dané tým, že v roku 2001 mobilní operátori Globtel a EuroTel postupne prestali prevádzkovať voľné SMS brány. 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, Centrum a Surf.

Najviac vyhľadávané termíny na niektorých slovenských vyhľadávačoch za rok 2001

	Atlas	Centrum	Surf
1.	SMS	SMS	sex
2.	mp3	sex	SMS
3.	sex	www	gay
4.	porno	porno	mp3
5.	mobil	práca	erotika
6.	mapy	mob	porno
7.	práca	CD	lolita
8.	pohľadnice	mobil	práca
9.	zoznamka	video	pohľadnice
10.	gay	chat	inzercia
11.	erotika	web	video
12.	inzercia	zamestnanie	nokia
13.	hry	zoznamka	mobil
14.	logo	brigáda	reality
15.	nokia	obchod	knihy
16.	chat	zoznámenie	chat
17.	práca	pokec	download
18.	auto	globtel	obrázky
19.	referáty	súťaž	hry
20.	zamestnanie	shop	zadarmo

O preferenciách používateľov slovenského internetu svedčí aj poradie návštevnosti stránok podľa informačného auditu služby NAJ.SK.

Poradie návštevnosti slovenských stránok za rok 2004 podľa NAJ.SK

Poradie	Počet prístupov	Stránka
1.	6 942 125	Zoznam Hlavná Stránka
2.	5 515 403	AZet.sk
3.	4 612 202	Sector – hry, programy, filmy, kultúra
4.	4 609 014	SME on-line – sme.sk
5.	3 677 705	POST.sk – Vaša osobná e-mail schránka zadarmo!
6.	3 559 819	Bleskovky
7.	3 411 111	SkyEurope Airlines
8.	3 071 167	POBOX.SK – e-mail zadarmo
9.	2 992 089	AZet.sk – Rýchla pošta
10.	2 056 689	Zoznam E-mail
11.	1 543 800	On-line hry
12.	1 426 901	PROFESIA – burza práce na Internete
13.	1 294 767	Funny.sk
14.	1 214 694	Centrum.sk – hlavná stránka
15.	1 194 342	LIMBA – dovolenka na Slovensku a v ČR
16.	1 082 010	AZet.sk – Email
17.	1 081 978	referaty.sk – Všetko, čo študent potrebuje
18.	1 076 864	SUPERMUSIC.SK
19.	1 052 754	Centrum.sk – mail
20.	1 016 073	AllianceWeb

V mesiacoch január až marec 2003 sme analyzovali transakčné logy na serveri SURF.SK. (<http://www.surf.sk>), ktorý obsahoval informácie o viac ako 20 000 webových stránkach slovenského internetu. Adresár je hodnotený, najlepšie stránky sú v kategórii TOP. Systém mal 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) apod.

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 nedodržiavajú nasledovné odporúčania informačných architektov týkajúce sa vyhľadávania na webových stránkach:

- Spojenie na rozšírené vyhľadávanie by mala byť umiestnené pod alebo vedľa okienka vyhľadávania.
- Na stránke rozšíreného vyhľadávania by mala byť linka na vysvetlenie, ako používať možnosti rozšíreného vyhľadávania. Táto linka by mala byť v blízkosti tlačidla „vyhľadať“.
- Ako doplnok k možnosti používania boolovských operátorov AND/OR v okienku rozšíreného vyhľadávania by používateľom mali byť k dispozícii voľby „všetky slová“, „akékoľvek slovo“ a „fráza“. Táto možnosť by mala prítomná vzhľadom na fakt, že väčšina používateľov nevie správne narábať s operátormi AND/OR/NOT.

V nasledujúcej tabuľke je uvedené, ako spĺňajú uvedené odporúčania najviac využívané slovenské vyhľadávacie nástroje Zoznam slovenského Internetu, Atlas.sk, Centrum.sk, Surf.sk a Superzoznam.

Ako spĺňajú slovenské vyhľadávacie nástroje odporúčania informačných architektov pre rozšírené vyhľadávanie

	Zoznam slovenského Internetu	Atlas.sk	Centrum.sk	Surf.sk	Superzoznam
Odporúčanie 1	nie je spojenie na rozšírené vyhľadávanie	spojenie na rozšírené vyhľadávanie je vedľa okienka na základné vyhľadávanie	spojenie na rozšírené vyhľadávanie je vedľa okienka na základné vyhľadávanie	nie je, rozšírené vyhľadávanie sa dá použiť už z úvodnej stránky	nie je
Odporúčanie 2	nie je	nie je	priamo na stránke rozšíreného vyhľadávania sú podrobne vysvetlené všetky možnosti	na úvodnej stránke je spojenie na detailné vysvetlenie možnosti systému	nie je
Odporúčanie 3	nie je	nie je	nie je	nie je	nie je

Ako vidíme z tabuľky, ani jeden zo slovenských nástrojov nespĺňa odporúčania informačných architektov, ktoré sú už skoro štandardom u svetových vyhľadávačov (Google, AlltheWeb, AltaVista, Teoma).

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 prúžkami nachádzajúcimi sa na navštevovaných www serveroch a motiváciou na vyplnenie dotazníka boli aj zaujímavé ceny, ktoré ponúkli sponzori súťaže. Prieskumu sa zúčastnilo 1676 respondentov. Respondenti odpovedali na 30 otázok. Otázky sa týkali používaných technológií a spôsobu pripájania, používaných vyhľadávacích nástrojov a samotného obsahu. Išlo o prvý prieskum tohto druhu na Slovensku. Niektoré výsledky prieskumu:

- Používatelia slovenského internetu využívajú vyhľadávacie nástroje, pretože až 78,04 % používateľov sa o nových stránkach dozvedá z vyhľadávačov.
- 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.
- 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 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.
- 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 %) a 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 počet výskumov správania sa používateľov narastá, ešte stále vieme iba veľmi málo o informačných potrebách používateľov pri vyhľadávaní informácií. V súčasnosti využíva internet skoro jedna miliarda používateľov a je zrejmé, že informačné potreby sú rozdielne, na aké sme boli zvyknutí pri vyhľadávaní v komerčných databázových centrách, kde väčšinou išlo o získanie informácií k predmetu výskumu, prípadne v digitálnych knižniciach.

Na túto skutočnosť poukazuje Andrei Broder (2002). Na základe výskumu transakčných logov v AltaViste zistil, že informačná potreba nemusí byť vždy iba informačná, ale môže byť aj navigačná (hľadáme URL niektorej stránky), prípadne transakčná (z internetu chceme niečo stiahnuť, kúpiť apod.). Informačné typy požiadaviek sa najviac približujú ku klasickým typom požiadaviek, na aké sme boli zvyknutí v databázových centrách. Čo je ale zaujímavé, väčšinou na našu požiadavku získame zoznam relevantných ďalších spojení a nie jeden relevantný dokument.

V nasledujúcej tabuľke vidíme percentuálne zastúpenie jednotlivých typov požiadaviek analýzou dotazníka a transakčných logov.

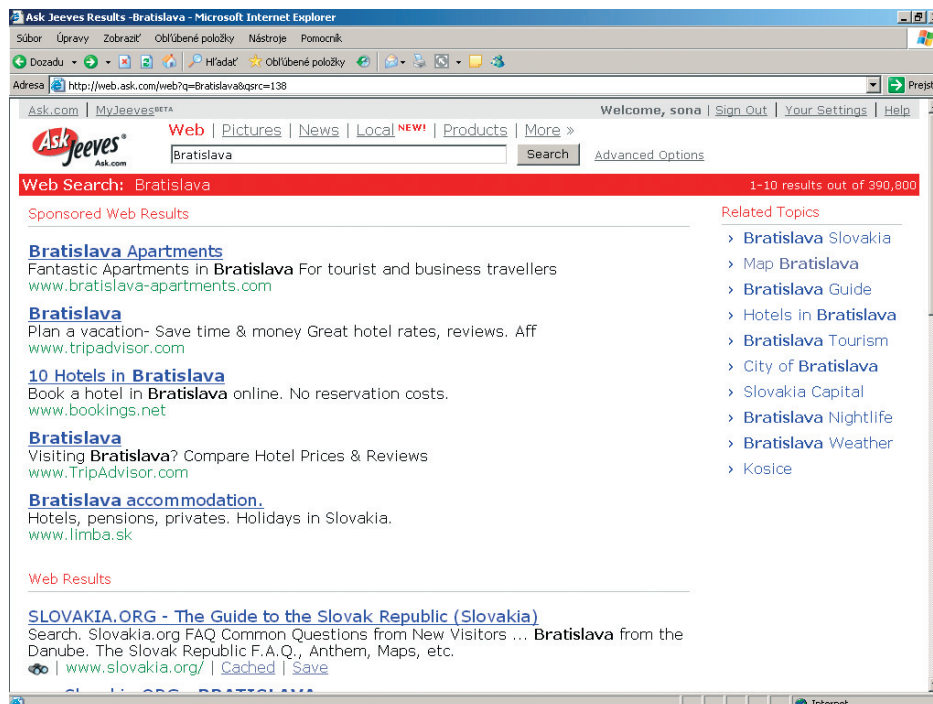
Analýza typov požiadaviek podľa Andrei Brodera

Typ požiadavky	Dotazník	Transakčné logy
Navigačná	24,5 %	20 %
Transakčná	36 %	30 %
Informačná	39 %	48 %

Na základe uvedenej taxonómie môžeme identifikovať aj tri generácie vyhľadávacích nástrojov:

Prvá generácia využívala predovšetkým text na stránkach a približovala sa klasickému informačnému prieskumu. Podporovala predovšetkým informačné požiadavky, typickými predstavitelmi prieskumových strojov boli AltaVista, Excite, WebCrawler v rokoch 1995 až 1997.

Vznik **druhej generácie** vyhľadávacích nástrojov sa datuje od roku 1998. Pri výpočte algoritmu sa využívali už aj texty, ktoré sa nenachádzali priamo na stránke, ako napríklad analýza spojení, text spojení, analýza počtu kliknutí apod. Podporovali sa informačné aj navigačné požiadavky, Google patril k prvým nástrojom, ktorý pri radení dokumentov využíval predovšetkým analýzu spojení, a DirectHit sa sústreďoval na analýzu počtu kliknutí. Vyhľadávacie nástroje **tretej generácie** už kombinujú údaje z viacerých zdrojov s cieľom zohľadniť informačnú potrebu, ktorá je za určitým typom požiadavky. Tak napríklad na požiadavku **Bratislava**, vyhľadávací nástroj nám ponúkne aj ďalšie súvisiace odkazy ako napríklad **počasie v Bratislave, ubytovanie v Bratislave, sprievodca po Bratislave, kultúrne podujatia v Bratislave** apod. Cieľom je podporovať všetky typy dotazov. Na obrázku je ukážka odpovede na požiadavku **Bratislava** personalizovanej verzie vyhľadávača **Ask Jeeves**.



Vyhľadávací nástroj tretej generácie Ask Jeeves

3.2 Problémy vyhľadávacích nástrojov

Problémy, ktoré musia riešiť projektanti vyhľadávacích nástrojov, by sme mohli zhrnúť nasledovne:

- množstvo dokumentov v internete,
- veľkosť a aktualizácia databázy súvisiaca s rýchlou zmenou obsahu a štruktúry spojení,
- spam,
- kvalita obsahu dokumentov v internete,
- hodnotenie kvality výsledkov vyhľadávania,
- neštruktúrované dáta.

V súčasnosti je jedným z najväznejších problémov internetu nájdenie kvalitnej a relevantnej informácie, ktorá zodpovedá našim informačným potrebám. Problém sa komplikuje tým, že množstvo informácií v digitálnej forme neustále narastá. Na School of Information Management & Systems University of California, Berkeley sa od roku 2000 rieši projekt, ktorého cieľom je zistiť, koľko informácií sa publikuje každý rok. V štúdiu Petra Lymana a Hala Variana **How Much Information 2003** sa uvádza, že dnes je stále viac informácií iba v digitálnej forme. Čo je ale neuveriteľné, tlačенý obsah reprezentuje iba 0,003% celého obsahu publikovaného vo svete. Znamená to, že každej jednej vete v tlačенých médiách zodpovedá 30 000 viet v digitálnej forme. Keď sa v minulosti hovorilo o informačnom preťažení čitateľov, dnes to platí niekoľkonásobne. V roku 2003 sa vyprodukovalo v digitálnej forme 5 exabytov novej informácie. Pre porovnanie, Kongresová knižnica v digitálnej forme by obsahovala asi 136 terabytov informácií, čiže v roku 2003 sa publikovalo množstvo informácií, ktoré zodpovedá 37 000 novým knižniciam o veľkosti Kongresovej knižnice.

K ďalším vážnym problémom patrí veľkosť a aktualizácia databázy súvisiaca s rýchlou zmenou obsahu a štruktúry spojení. Podľa Nielsen/NetRatings používatelia webu strávia mesačne iba na Googli 13 miliónov hodín. Prieskumové stroje majú vlastný program (spider, robot), ktorý zindexuje jednotlivé stránky a obsah si uloží do svojho indexu, aby mohol rýchlo nájsť odpoveď. Vzhľadom na to, že stránky sa často menia, prieskumové stroje sú nútené často aktualizovať svoj index. To bol tiež jeden z dôvodov výskumného projektu **Ako sa vyvíja web z pohľadu prieskumových strojov** (Ntoulas, Cho, Olston, 2004). Výskum sa zamerl na analýzu zmien v štruktúre spojení a v obsahu webových stránok na základe monitorovania 154 veľkých webových sídiel každý týždeň počas obdobia jedného roka. Išlo o prvý prieskum, ktorý analyzoval zmeny v štruktúre hypertextových spojení webových sídiel. Význam výskumu spočíval v tom, že algoritmus výpočtu relevancie nájdených záznamov prieskumových strojov berie do úvahy okrem obsahu stránky aj význam a dôležitosť spojení, ktoré vedú na stránku. Napríklad Google používa patentovanú technológiu **Page Rank**, ktorá skúma okrem obsahu aj aké spojenia vedú na stránku a podľa toho dáva stránke dôležitosť.

Výsledky výskumu poukazujú na niektoré zaujímavé skutočnosti:

- Autori štúdie odhadujú, že každý týždeň pribudne na webe asi 8% nových stránok. Podľa veľkosti Google by mal mať súčasný povrchový web asi 8 miliárd stránok, čo znamená, že každý týždeň pribudne asi 640 miliónov nových stránok.
- Ďalej sa predpokladá, že po roku je na webe asi 50% úplne nového obsahu.
- Každý týždeň sa pridáva asi 25% nových spojení. Po roku je asi 80% spojení nahradených novými spojeniami. Štruktúra spojení sa mení ešte rýchlejšie ako obsah, čo znamená prehodnotiť stratégiu radenia relevantných dokumentov na výstupe.

- Existujúce stránky na webe sa menia pomerne málo, ale často sú nahrádzané úplne novými stránkami. Tieto väčšinou preberajú obsah od starých stránok (redesign stránok, adaptácia na nové technológie).
- Životnosť stránok je pomerne nízka, a preto je veľmi dôležité vytvárať archív webu, príkladom môže byť celosvetový projekt Internet Archive (<http://www.archive.org>), v Čechách WebArchive ako digitálna knižnica českého webu (<http://webarchiv.nkp.cz>).

Závažný problém, ktorý treba brať do úvahy, je spam vyhľadávacích nástrojov. Spam spočíva v snahe tvorcov webových stránok využiť nekalé techniky na to, aby stránka dosiahla čo najvyššie umiestnenie pre dané kľúčové slová. Je to vážny problém, pretože podľa výskumov 85 % používateľov si pozrie záznamy iba na prvej strane obrazovky a počet spamov narastá. Často sa v prvej desiatke objavia záznamy, ktoré vôbec nekorešpondujú s obsahom kľúčových slov. Vyhľadávacie nástroje už majú definovanú stratégiu boja proti spamu a takéto stránky jednoducho vylúčia zo svojho indexu. Prehľad rôznych nástrojov a techník spamu je uvedený v článku Web Spam Taxonomy (Gyongyi – Garcia-Molina, 2004). V súčasnosti už aj u nás vznikajú mnohé služby ponúkajúce optimalizáciu stránok pre vyhľadávacie nástroje, tzv. SEO – Search Engine Optimization. Znamená to vytvorenie, alebo úpravu webových stránok tak, aby dosahovali popredné miesta vo vyhľadávačoch na dané kľúčové slová. Je diskutabilné, čo sa v tomto prípade už považuje za spam.

Kvalita obsahu dokumentov v internete je daň, ktorú platíme za demokratický charakter internetu. To je tiež jeden z dôvodov, prečo niektoré algoritmy radenia dokumentov na výstupe rešerše berú do úvahy tiež počet a kvalitu hypertextových spojení, ktoré vedú na daný dokument. Ukazuje sa potreba venovať väčšiu pozornosť výskumu správania sa používateľov pri hodnotení výsledkov rešerše. Aj keď používatelia nedávajú explicitne spätnú väzbu, veľa sa dá zistiť z analýzy prezretých dokumentov. Stále väčší význam majú kontrolované predmetové brány internetu, kde pri hodnotení dokumentov vstupuje aj ľudský faktor.

K ďalšej nevýhode patria neštruktúrované dáta, ktoré sú dané jazykom HTML. Jazyk HTML ako značkovací jazyk sa zaoberá prezentačnou funkciou WWW dokumentov a iba veľmi málo hovorí o sémantike dát. Tento jazyk popisuje štruktúru webovej stránky a dáva webovému prehliadaču informáciu o tom, ako má stránku vykresliť: čo má byť napísané veľkými písmenami, čo malými, kde má byť ktorá časť textu zobrazená atď. Následníkom jazyka HTML je Extensible Mark-up Language (XML). Poskytuje oproti HTML rad výhod, hlavný rozdiel spočíva v oddelení obsahu a prezentácie.

Na XML nadväzuje množstvo výskumných a vývojových aktivít. K významnej patrí Resource Description Framework (RDF), ktorý umožňuje kódovanie, výmenu a opakované využitie štruktúrovaných metaúdajov, ktoré je založené na využití XML ako výmennej syntaxe. Týmto spôsobom sa podporuje integrácia rôznych aplikácií.

4. Navrhované riešenia

O tom, že web predstavuje neoceniteľný zdroj informácií, dnes nikto nepochybuje. Vyhľadávač Google predstavil beta verziu novej služby **Google Print** (print.google.com), v rámci ktorej bude veľká časť textov uložených v najvýznamnejších knižniciach na svete prevedená do digitálnej podoby. Vďaka novému projektu získajú používatelia internetu prístup k plným textom mnohých kníh, staršie knihy, ktorým už vypršala ochrana autorskými právami, budú dostupné kompletne. Pri ďalších sa zobrazia len úryvok obsahujúci žiadané vyhľadávané slovo alebo frázu, ktorý bude doplnený informáciou o tom, kto knihu vydal, kde si ju možno kúpiť, prípadne v ktorej

najbližšej knižnici sa dá požičať. Spolu bude zdigitalizovaných viac ako pätnásť miliónov kníh. Podobne beta verzia novej služby **Google Scholar** (scholars.google.com) umožňuje vyhľadávať vedecké články, dizertácie, výskumné správy. Zahrnutím citácií do výsledkov vyhľadávania nová služba zároveň poskytuje aj zoznam ďalších textov k skúmanej téme. Čiže máme k dispozícii množstvo informácií v digitálnej forme, ktoré dokážu počítače prečítať, ale nedokážu mu porozumieť. Výsledkom je potom nespokojnosť používateľov. Kvalita výsledkov je determinovaná nedokonalosťou vyhľadávacích možností na webe, zväčša založených na princípe formálnej zhody medzi rešeršnou požiadavkou a výsledkom prieskumu a na fulltextovom prehliadaní neumožňujúcim formulovať sémanticky komplexné požiadavky.

Tieto problémy sú v hlavnej miere 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. **Sémantický web** by túto situáciu mal zmeniť, dať obsahu strojom zrozumiteľnú formu umožňujúcu počítačom a iným zariadeniam informácie hľadať, využiť a takto registrovať a riešiť aj komplexné úlohy, ktorých riešenie je dnes vyhradené len pre ľudí.

Riešenie problému na globálnej úrovni je v **sémantickom webe**. Sémantický web je rozšírením dnešného webu, v ktorom bude daný informáciám jasne definovaný význam, lepšie umožňujúci kooperáciu počítačov a ľudí (Berners-Lee – Hendler – Lasilla, 2001). Uvedenú problematiku rozpracováva predovšetkým World Wide Web konzorcium v rámci aktivity **Sémantický web** (<http://www.w3.org/2001/sw/>).

Už čoskoro bude nevyhnutné, aby obrovské množstvo cez web prístupných dokumentov a dát bolo popísané takým spôsobom, ktorý umožní počítačom porozumieť ich obsahu a spracovať ich. Toto by umožnilo nasadenie tzv. agentov, čiže strojov, ktoré by z web stránok zbierali údaje o ich obsahu a vymieňali by si ich s údajmi od iných agentov. Sémantický web umožní, aby inteligentní agenti pre nás vykonávali celý rad zložitých úloh. Jediný rozdiel oproti doterajšiemu webu bude v tom, že každá informácia bude mať presne definovaný význam, čo umožní oveľa lepšiu kooperáciu, vyhľadávanie a spracovanie informácií (Makulová 2002, s. 337–338.).

Najdôležitejšie koncepcie, technológie, protokoly a štandardy na rozvoj sémantického webu sú už k dispozícii. V kontexte sémantického webu majú veľký význam **ontológie**, ktoré môžeme charakterizovať ako formálny opis objektu a jeho vzťahov s okolím. Ich cieľom je umožniť strojom komunikovať navzájom celkom bez alebo len s obmedzeným zásahom človeka. Na ich modelovanie sú dnes dostupné tzv. topic maps jako nový ISO štandard na opis poznatkových štruktúr a ich previazanie na informačné pramene.

Pri vyhľadávaní v internete rozlíšujeme predovšetkým dva druhy aktivít, a to:

- vyhľadávanie v celom internete,
- vyhľadávanie v rámci jednotlivého sídla.

Preto je potrebné riešiť aj problematiku na dvoch úrovniach.

4.1 Vyhľadávanie v celom internete

Vyhľadávanie v celom internete umožňujú predovšetkým vyhľadávacie nástroje internetu, o ktorých sme hovorili v úvode. Problém rieši predovšetkým sémantický web a technológie XML, RDF, OML apod. V tejto súvislosti hrá mimoriadne dôležitú úlohu tvorba kvalitných dokumentov, ktoré zohľadňujú webové štandardy. Uvedená problematika je veľmi pekne a prehľadne spracovaná v príspěvku **Kvalitní dokument jako základ účinného vyhledávání informací** (Tkačiková, 2004).

Stále väčšie množstvo digitálnych informácií vyžaduje riešiť otázky **interoperability**.

Interoperabilita digitálneho obsahu predstavuje jeho opakovateľnú využiteľnosť, prenášateľnosť apod. v rôznych sieťach, systémoch a organizáciách. Kľúčom na jej dosiahnutie sú normy – dohodnuté pravidlá a smernice na tvorbu, opis a správu digitálneho obsahu (Gill – Miller, 2002).

Čiastočným riešením množstva vyhledaných informácií prezentovaných používateľovi riešia personalizované vyhľadávacie nástroje, ktoré si používateľ prispôsobuje svojim informačným potrebám. Výskum v tejto oblasti umožňuje radiť nájdené záznamy na základe osobných preferencií používateľa. Napokon je to tvorba kvalitných predmetových brán, ktoré cieľavedome budujú informační špecialisti, ako napríklad Resource Discovery Network, Infomine apod. K výborným pomôckam patrí príručka na budovanie kvalitných portálov na internete **DESIRE Information Gateways Handbook** (2000).

4.2 Vyhľadávanie v rámci jednotlivého sídla

Situácia je ďaleko komplikovanejšia pri vyhľadávaní v rámci jednotlivých sídiel, na ktoré sme nasmerovaní z jednotlivých vyhľadávačov. Iba málo sídiel má nainštalované vyhľadávanie a organizácia obsahu nezodpovedá požiadavkám informačnej architektúry sídla. Podľa ostatného prieskumu **Search Tools** z júla 2002 z 1075 sídiel malo vyhľadávanie nainštalované iba 33 % sídiel. Príčiny, ktoré viedli k inštalácii vyhľadávania, boli predovšetkým zlepšenie navigácie (57 %), profesionálny vzhľad sídla (23,7 %), služby zákazníkom (6,6 %), marketing (4,7 %), užitočný nástroj sídla (2,5 %) a iné dôvody v 5,4 % (Why Site Managers Install Search Engines). Príčiny, pre ktoré na stránke nebolo nainštalované vyhľadávanie, boli predovšetkým nedostatok času (42,3 %), príliš komplikované (15,1 %), vysoká cena (9,8 %), nemožnosť používať CGI (6,4 %), málo obsahu (6,1 %), nové sídlo (4 %), nie je potreba (3,8 %), v databáze sa dá vyhľadávať (2,8 %), hľadanie vhodného nástroja (2,4 %), neviem (2,4 %) a iné dôvody 4,9 % (Reasons Why Search is Not Installed).

V rámci jednotlivých sídiel by sa pri dizajne a redizajne mala uplatňovať metodológia **informačnej architektúry sídla**. Informačná architektúra vznikla ako odpoveď na nárast informácií v digitálnej forme a jej cieľom je nájsť odpoveď na otázku, akým spôsobom je možné zorganizovať také veľké množstvo obsahu v digitálnom priestore tak, aby používatelia dokázali nájsť informáciu a orientovať sa v ňom.

Existuje viac prístupov a chápaní pojmu informačná architektúra, k jednej z najkomplexnejších patrí definícia Louisa Rosenfelda a Petra Morvilla (2002), ktorí ju charakterizujú nasledovne:

- kombinácia organizačnej, navigačnej a popisnej schémy informačného systému,
- štruktúrny dizajn informačného priestoru s cieľom uľahčiť vykonávanie úloh a umožniť intuitívny prístup k obsahu,
- umenie a veda štruktúrovania a klasifikovania webových sídiel a intranetu tak, aby ľudia dokázali nájsť a riadiť informácie,
- vznikajúca disciplína a spoločenská prax s cieľom zaviesť princípy dizajnu a architektúry do digitálneho sveta.

Dôležité je uvedomiť si, že informačná architektúra je veda, ale zároveň aj umenie štruktúrovať a klasifikovať informácie na webe tak, aby ľudia dokázali nájsť a riadiť informácie. Vyplýva z toho, že návrh informačnej architektúry síce vychádza z overenej metodológie, ale vždy záleží od intuície a skúsenosti informačného architekta, akým spôsobom zorganizuje informácie

v digitálnom priestore. Pri dnešnom enormnom náraste webových sídiel platí viac ako inokedy: informáciu, ktorú nikto nenájde, ani nikto neprečíta. Pritom je jasné, a presvedčujú nás o tom aj výsledky prieskumov, že ľudia prichádzajú na web hlavne s cieľom nájsť informácie.

Užší prístup k informačnej architektúre je v prístupe Gerry McGoverna a Roba Nortona (2002). Podľa nich je informačná architektúra vedná disciplína, ktorá sa zaoberá organizáciou a rozvrhnutím obsahu na webovom sídle. Špeciálne sa zaoberá vývojom metaprvkov, klasifikáciou, navigáciou, vyhľadávaním a rozvrhnutím obsahu.

My sa prikláňame skôr k širšej definícii, pretože informačný architekt musí úzko spolupracovať s viacerými disciplínami, ako napr. s grafickým dizajnom, usability inžinierstvom, softvérovým inžinierstvom, manažmentom obsahu, manažmentom znalostí apod. Informačná architektúra ako vedná disciplína sa vyučuje už na viacerých katedrách knižničnej a informačnej vedy. Podobne na Katedre knižničnej a informačnej vedy Filozofickej fakulty Univerzity Komenského v novom učebnom akreditovanom programe od roku 2005 bude patriť informačná architektúra ku kľúčovým disciplinám odboru.

Pri tvorbe informačnej architektúry má veľký význam tvorba **metadát**. Termín metadáta je výstižný na zvýraznenie veľkého rozdielu v udržiavaní katalógov kníh v tradičných knižniciach a aktivitami, ktoré sa realizujú v záujme poskytnutia týchto katalógov prostredníctvom internetu. Odráža to zmenu v katalogizácii, ktorá by sa dala charakterizovať aj ako katalogizácia digitálnym smerom. Metadáta používame na pomenovanie informácií o nejakom zdroji, ktorý nám umožní identifikovať, lokalizovať a žiadať daný informačný zdroj.

Metadáta môžu byť: opisné dáta (autor, titul...), subjektové dáta (kľúčové slová, opisy...), prístupové dáta (opis HW a SW požiadaviek na použitie zdroja), administratívne dáta (opis vlastných metadát – kedy a kým boli vytvorené...), informácie o pravidlách a podmienkach použitia (Ukropová-Strapcová, 2001). Podľa nášho názoru je to predovšetkým tvorba metadát, kde môžu knižniční a informační pracovníci významne pomôcť a využiť svoje know-how z organizácie poznania. Svedčí o tom aj množstvo úspešných projektov, z ktorých k najvýznamnejším patrí Dublin Core. Začlenením Dublin Core metadát do hlavičky HTML je možné zlepšiť presnosť vo vyhľadávaní.

4.3 Výskumy správania používateľov

Pri riešení problémov zefektívnenia vyhľadávania informácií v internete je potrebné orientovať sa na nové smery vo výskume správania sa používateľov. Predchádzajúce práce venované výskumu správania používateľov vo webovskom priestore sa orientovali skôr na to, **ako ľudia vyhľadávajú, ako vyjadrujú svoju informačnú potrebu, čo hľadajú a nie prečo tak hľadajú**. Čiže do úvahy sa brali také faktory, ako je typická dĺžka požiadavky, pomocou koľkých slov sa skladá typická požiadavka, ako sa používajú výroky Boolovej algebry apod. Výskumníci sa nesnažili sa pochopiť, prečo používateľ hľadá týmto spôsobom, aký je zámer, aká informačná potreba sa skrýva za požiadavkou.

Pritom je jasné, že odpoveď na otázku, **prečo používateľ hľadá**, je veľmi dôležitá, aby sme dokázali uspokojiť jeho informačnú potrebu. Používateľ vždy vyhľadáva s určitým cieľom, buď chce kúpiť darček na narodeniny svojmu blízkemu, prípadne chce zistiť, aké bude počasie v mieste destinácie jeho dovolenky, alebo potrebuje informácie k napísaniu seminárnej práce a pod. V niektorých prípadoch jedna a tá istá požiadavka môže viesť k uspokojeniu rôznych cieľov. Napríklad keramika – chcem kúpiť darček, hľadáme knihu o keramike apod. V prípade, že vyhľadávací nástroj pochopí zámer používateľa, môže prispôbiť interfejs jeho potrebám. Napríklad zobrazenie sponzorovaných reklamných odkazov môže byť relevantné v prípade, že

používateľ chce niečo kúpiť, ale nie ak robí rešerš k výskumnej práci. Podobne sa môže líšiť podľa zámeru používateľa aj algoritmus výpočtu relevantných záznamov. Čiže lepšie poznanie informačných potrieb používateľov vo významnej miere môže zlepšiť výkonnosť prieskumových strojov. Ako prvý poukázal na rôzne informačné potreby vo webe Andrei Broder (2002) a jeho výskum potvrdili a obohatili Rose a Levinson (2004).

4.4 Výchova používateľov

Ukazuje sa nevyhnutnosť permanentnej výchovy používateľov informácií na všetkých stupňoch vzdelávania. Na Katedre knižničnej a informačnej vedy bola v roku 2003 obhájená diplomová práca **Výchova k mediálnej gramotnosti** (Weisenbacher, 2003).

Na základe analýzy rôznych definícií a prístupov autor definuje mediálnu gramotnosť ako súbor vedomostí a znalostí umožňujúcu vedomé využívanie, analyzovanie a hodnotenie rôznych médií a ich obsahov. Potom výchovu k mediálnej gramotnosti chápeme ako cieľavedomú činnosť poskytujúcu subjektu možnosť získať vedomosti a znalosti vedúce k vytvoreniu mediálnej gramotnosti (Weisenbacher, 2003, s. 19). V práci je návrh koncepcie výchovy k mediálnej gramotnosti v podmienkach Slovenskej republiky pre základné, stredné a vysoké školy zároveň so základnými okruhmi vhodných tém a metódami realizácie, ako aj načrtnutím úlohy knižníc v programe výchovy k mediálnej gramotnosti. V prípade dospelých jednotlivcov by sa počítalo s prednáškami, seminármi a kurzami, ktoré by organizovali verejné a vedecké knižnice pod metodickým vedením odborníkov na mediálnu gramotnosť, prípadne knihovníkov, ktorí by sa špecializovali na túto oblasť.

5. Záver

Cieľom príspevku bolo načrtnúť niektoré problémy, pred ktorými stoja používatelia pri vyhľadávaní informácií v internete. Ide o veľmi široký okruh interdisciplinárnych problémov, pred ktorými stoja samotní používatelia internetu ale aj tvorcovia vyhľadávacích nástrojov.

V súčasnej dobe prudkého nárastu digitálnych informácií stoja pred knižnicami ako aj odborníkmi z knižničnej a informačnej vedy nové úlohy. Knižnice pôsobia v sieťovom informačnom prostredí, rozvoj internetu umožňuje prístup k stále narastajúcemu objemu digitálnych informácií rozptýlených v internete aj z tej najmenšej knižničnej pobočky a preto by pracovníci knižníc mali vedieť čo najlepšie pomáhať svojim čitateľom pri objavovaní a vyhľadávaní informácií. Znalostná spoločnosť, celoživotné vzdelávanie a silnejúci dôraz na elektronickú komunikáciu s inštitúciami rôzneho druhu prináša nárast požiadaviek na jednoduchý prístup k informáciám pre všetkých občanov. Mnohé knižnice, ktoré vytvárajú a sprostredkujú digitálny obsah, ho musia správne popísať tak, aby sa dal používateľom ľahko vyhľadať a bol interoperabilný s ostatným digitálnym obsahom.

Je dôležité, aby sme si uvedomili, že obrovské množstvo potenciálnych informácií v internete neznamená, že budeme múdrejší. Kapacita ľudskej pamäte je obmedzená, a preto záleží od samotných používateľov, ako ich dokážu interpretovať, hodnotiť a integrovať do svojho osobnostného fondu.

Budovanie učiacej sa spoločnosti neznamená sústrediť sa iba na technológiu, ale na obsah a ľudí, ako tvorcov a zároveň aj používateľov informácií. Všetkým by nám malo ísť o to, aby sme dokázali uspokojiť ich informačné potreby na úrovni tretieho tisícročia.

Literatúra:

- BERGMAN, M. K. 2001. The Deep Web: Surfacing Hidden Value. [cit. 2004-12-10]. Dostupné na internete: <http://www.brightplanet.com/technology/deepweb.asp>
- BERNERS-LEE, T., HENDLER, J., LASSILA, O. 2001. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. In *Scientific American*, 2001, vol. 284. [cit. 2004-12-23]. Dostupné na internete: <http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2>
- BRODER, A. 2002. A taxonomy of Web Search [on-line]. *ACM Sigir Forum*, 2002, vol. 36, no. 2, 2002 [cit. 2004-12-10]. Dostupné na internete: <http://doi.acm.org/10.1145/792550.792552>
- BROOKS, Terrence A. 2003. Web Search: how the Web has changed information retrieval. [on-line] In *Information Research*, 2003, vol. 8, no. 3, paper no. 154, [cit. 2004-12-10]. Dostupné na internete: <http://InformationR.net/ir/8-3/paper154.html>
- eEurope+2003: progress report: february 2004. European Commission, 2004, 49p.
- DESIRE Information Gateways Handbook. [cit. 2005-01-13]. Dostupné na internete: <http://www.desire.org/handbook/welcome.html>
- GILL, T., MILLER, P. 2002. Re-inventing the Wheel? Standards, Interoperability and Digital Cultural Content. In *D-Lib Magazine* 2002, vol. 8, no. 1 [cit. 2005-01-13]. Dostupné na internete: <http://www.dlib.org/dlib/january02/gill/01gill.html>, ISSN 1082-9873.
- GYONGYI, H., GARCIA-MOLINA, H. 2004. Web Spam Taxonomy. Technical Report [on-line], Stanford University, March 2004. [cit. 2004-12-10]. Dostupné na internete: <http://dbpubs.stanford.edu:8090/pub/showDoc.Fulltext?lang=en&doc=2004-25&format=pdf&compression=&name=2004-25.pdf>
- JANSEN, B. J., POOCH, U. 2000. Web user studies: A review and framework for future work. In *Journal of the American Society for Information Science and Technology*, 2000, vol. 52, no. 3, p. 235–246.
- JANSEN, BERNARD, J. 2000. The effect of query complexity on Web searching results. In *Information Research*, 2000, vol. 6, no.1 [cit. 2003-02-20]. Dostupné na internete: <http://www.shef.ac.uk/~is/publications/infres/paper87.html>
- LARSEN, R. L. 1997. Relaxing assumptions: stretching the vision. In *D-Lib Magazine* 1997, vol. 3, no. 4 [cit. 2005-01-10]. Dostupné na internete: <http://www.dlib.org/dlib/april97/04larsen.html>
- LYMAN, P., VARIAN, H. R. 2004. How Much Information 2003. [cit. 2005-01-10]. Dostupné na internete: <http://www.sims.berkeley.edu/how-much-info>
- MAKULOVÁ, Soňa 2003. Informačné správanie používateľov pri vyhľadávaní informácií v internete. In *Informačné správanie a digitálne knižnice*. Bratislava: CVTI SR, 2003, s. 90–110.
- MAKULOVÁ, Soňa. 2002. Vyhľadávanie informácií v internete: problémy, východiská, postupy. Bratislava: EL&T, 2002. 376 s. ISBN 80-88812-16-X.
- MCGOVERN, G., NORTON, R. 2002. *Content Critical*. London: Pearson Education Limited, 2002, 241 s.
- NTOULAS, A., CHO, J., OLSTON, C. 2004. What's New on the Web? The Evolution of the Web from a Search Engine Perspective. In *Proceedings of the Thirteenth WWW Conference*. New York, USA: 2004. [cit. 2004-12-10]. Dostupné na internete: http://oak.cs.ucla.edu/~ntoulas/pubs/ntoulas_new.pdf
- PAPÍK, R. 2001. Vyhľadávaní informácií I. Umění či věda? In *Národní knihovna*, 2001, roč. 12, č.1, s. 18–25. [cit. 2005-01-10]. Dostupné na internete: <http://full.nkp.cz/nkkr/NKKR0101/0101018.html>
- PAPÍK, R. 2001. Vyhľadávaní informácií II. Uživatelské rozhraní a vlivy oboru "human-computer interaction". In *Národní knihovna*, 2001, roč. 12, č. 2, s. 81–90. [cit. 2005-01-10]. Dostupné na internete: <http://full.nkp.cz/nkkr/NKKR0102/0102081.html>
- Reasons Why Search is Not Installed. *SearchTools Survey – July 2002*. [cit. 2005-01-12]. Dostupné na internete: <http://www.searchtools.com/surveys/survey05/reasons-not.html>
- ROSE, D. E., LEVINSON, D. 2004. Understanding user goals in web search. In *Proceedings of the 13th international conference on World Wide Web*. New York: ACM Press, 2004, s. 13–19.

- ROSENFELD, L., MORVILLE, P. 2002. *Information Architecture for the World Wide Web*. 2. vyd. Sebastopol: O'Reilly&Associates, 2002, 461 s. ISBN 0-596-00035-9
- SHERMAN, Ch., PRICE, G. 2001. *The Invisible Web*. Medford: Information Today, Inc., 2001. 439 s.
- SPINK, A., WOLFRAM, D., JANSEN, M. B. J., SARACEVIC, T. 2001. Searching the Web: the public and their queries. In *Journal of the American Society for Information Science and Technology*, vol. 52, no. 3, p. 226–234.
- SPINK, A., JANSEN, B.J., WOLFRAM, D., SARACEVIC, T. 2002. From e-sex to e-commerce: Web search changes. In *IEEE Computer*, vol. 35, no. 3, p. 133–135.
- 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*, vol. 36, no. 2, s. 207–227.
- MAKULOVÁ, Soňa. 2002. *Vyhľadávanie informácií v internete: problémy, východiská, postupy*. Bratislava: EL&T, 2002. 376 s. ISBN 80-88812-16-X
- The 2004 e-readiness rankings: a white paper from the Economist Intelligence Unit [on-line]. London; New York; Hong Kong: EIU, 2004 [cit. 2004-12-07]. Požaduje sa: Acrobat Reader. Dostupné na internete: http://graphics.eiu.com/files/ad_pdfs/ERR2004.pdf
- UKROPOVÁ, D., STRAPCOVÁ, E. 2001. Metadáta a čo s nimi. In *INFOS 2001*. Bratislava: Spolok slovenských knihovníkov, 2001. [cit. 2005-01-12]. Dostupné na internete: <http://www.aib.sk/infos/infos2001/34.htm>
- TKÁČIKOVÁ, D. 2004. Kvalitní dokument jako základ účinného vyhledávání informací. In *Informace na dlani 2004. Inforum 2004: 10. konference o profesionálních informačních zdrojích 25.–27. května 2004*, Vysoká škola ekonomická Praha.[CD-ROM]. Praha: Albertina icome, 2004. ISSN 1214-1429
- TOPOL, Jakub. *Informatizácia: pred SR aj Bulharsko a Rumunsko* [2004-06-25 11:33:28]. <http://www.lavica.sk/index.php?dzial=sk&id=144>
- WEISENBACHER, Peter. 2003. *Výchova k mediálnej gramotnosti*. Diplomová práca. Bratislava: Filozofická fakulta, 2003, 75 s.
- Why Site Managers Install Search Engines. SearchTools Survey – July 2002. [cit. 2005-01-12]. Dostupné na internete: <http://www.searchtools.com/surveys/survey05/reasons.html>

Publikácia bola spracovaná v rámci vedeckého grantového projektu VEGA 1/2481/05 Využívanie informácií pri informačnom správaní vo vzdelávaní a vede.