

ODBORNÁ ŠKOLENÍ A VZDĚLÁVÁNÍ PRACOVNÍKŮ ÚZEMNÍ VEŘEJNÉ SPRÁVY PRO OBLAST CESTOVNÍHO RUCHU

INFORMAČNÍ ZDROJE V CESTOVNÍM
RUCHU PRO VEŘEJNOU SPRÁVU

www.vzdelavanivcr.cz



INFORMAČNÍ ZDROJE V CESTOVNÍM RUCHU PRO VEŘEJNOU SPRÁVU

Autoři:

Mgr. Marek Waldhans, Ing. Radek Zasl

THEMA IT



Praha 2007

Památková péče, cestovní ruch a veřejná správa

Vydalo: Ministerstvo pro místní rozvoj ČR, Praha, 2007.

Staroměstské náměstí 6, 110 15 Praha 1, www.mmr.cz

Tato skripta byla vytvořena pro projekt „Odborná školení a vzdělávání pracovníků územní veřejné správy pro oblast cestovního ruchu“ CZ.04.1.03/4.2.00.1/0002 Operační program Rozvoj lidských zdrojů (OP RLZ), Opatření 4.2., Specifické vzdělávání.

Tento vzdělávací program je spolufinancován Evropským sociálním fondem (ESF) a státním rozpočtem ČR.

Obsah

ÚVOD	9
1 MODUL 1 - ZÁKLADNÍ INFORMAČNÍ ZDROJE V CESTOVNÍM RUCHU	11
2 MODUL 2 - OBECNÉ ZPŮSOBY UKLÁDÁNÍ DAT	14
2.1 STRATEGIE UKLÁDÁNÍ DAT	14
2.2 DATOVÁ MÉDIA	14
2.2.1 Dělení datových médií	15
2.3 KOMPRESCE DAT	16
2.3.1 Kompresní poměr	16
2.4 DATABÁZE	17
3 MODUL 3 - MODERNÍ TRENDY VE VYHLEDÁVÁNÍ NA INTERNETU	18
3.1 INTERNET A JEHO SLUŽBY	18
3.1.1 Historie internetu	19
3.1.2 Základní služby (protokoly)	19
3.1.3 Druhy a formáty internetového obsahu	21
3.1.4 Dostupnost internetového obsahu	23
3.2 VYHLEDÁVAČE	24
3.2.1 Vyhledávač/katalog/meta vyhledávač/Open Directory Project	24
3.2.2 Jak funguje vyhledávač	25
3.3 POKROČILÉ VYHLEDÁVÁNÍ	32
3.3.1 Vyhledávání ve fulltextovém vyhledávači	32
3.3.2 Vyhledávání v katalogu	39
3.3.3 Vyhledávání obrázků	41
3.3.4 Vyhledávání videa	43
3.3.5 Trendy ve vyhledávání	43
3.3.6 Mapy	45
3.4 OPTIMALIZACE PRO VYHLEDÁVAČE	47
3.4.1 Metody SEO	47
3.4.2 Etické metody	47
3.4.3 Neetické metody SEO, spam	48
3.4.4 Google bomba	49
3.5 PRÁVNÍ ASPEKTY VYHLEDÁVAČŮ	50
4 MODUL 4 - ZPŮSOBY VYHLEDÁVÁNÍ V INTERNÍCH A KOMBINOVANÝCH DATOVÝCH ÚLOŽIŠTÍCH	54
4.1 ULOŽENÍ INFORMACÍ	54
4.2 VLASTNOSTI APLIKACÍ PRO KOMPLEXNÍ VYTĚŽOVÁNÍ INFORMACÍ	57
4.3 ZÁKLADNÍ SCHÉMA ARCHITEKTURY	58
4.4 SITUACE NA TRHU	59
4.5 REPREZENTATIVNÍ SYSTÉMY PRO VYHLEDÁVÁNÍ DAT UVNITŘ ORGANIZACÍ	60
4.6 GOOGLE DESKTOP	60
4.6.1 Základy používání aplikace Google Desktop	61
4.6.2 Výsledky hledání	63
4.6.3 Indexování a správa aplikace	63
4.6.4 Pokročilejší hledání	64
4.6.5 Řešení pro organizace	66

4.6.6	<i>Další možnosti rozšíření.....</i>	67
4.6.7	<i>exalead one:desktop</i>	68
4.6.8	<i>Základy používání aplikace exalead one:desktop</i>	68
4.6.9	<i>Výrazy pro hledání</i>	69
4.6.10	<i>Uspořádání výsledku</i>	72
4.6.11	<i>Hledání ve výsledku.....</i>	73
4.6.12	<i>Vyhledávání ve vlastním PC a na Internetu současně</i>	74
4.6.13	<i>Řešení pro organizace</i>	76
4.7	POUŽITÁ LITERATURA	76
REJSTŘÍK POJMŮ.....		77

ÚVOD

Cílem tohoto vzdělávacího programu je vyškolit posluchače v efektivním způsobu vyhledávání informací vztahujících se k problematice cestovního ruchu i informací souvisejících. Tyto informace jsou uloženy v heterogenních datových úložištích mnoha státních i soukromých institucí a jsou často obtížně dostupné. Efektivně získávat informace je možné pouze tehdy, naučíme – li se používat moderní vyhledávací nástroje, které umožňují propojení různorodých zdrojů dat do jednotného informačního prostředí. Toto prostředí pak dovoluje vyhledávat a spojovat data jak ve vlastní organizaci a jejích datových úložištích, tak v prostředí virtuálních privátních sítí zúčastněných institucí i ve veřejných datových zdrojích. Systémy tohoto druhu umožní též kladení dotazů ve více jazycích, což má význam zejména v příhraničních oblastech. Významnou součástí je i vyhledávání údajů v multimediálním prostředí. K využití takových systémů je třeba určitých specifických znalostí, se kterými kurz posluchače seznámí.

Cílovou skupinou jsou zaměstnanci krajských, obecních a městských úřadů, zaměstnanci ministerstev, osoby pracující a podnikající v cestovním ruchu.

1 MODUL 1 - ZÁKLADNÍ INFORMAČNÍ ZDROJE V CESTOVNÍM RUCHU

V tomto modulu předkládáme souhrn nejvýznamnějších internetových portálů, obsahujících potřebné informace o cestovním ruchu.

<http://portal.env.cz/> **Portál životního prostředí** (topografické podklady, chráněná území, evropská soustava Natura 2000, povodně 2002, oblasti znečištění, historické mapování, přírodní, sociální a ekonomické indikátory udržitelného vývoje, Lisabonské indikátory).

<http://www.crr.cz/> **Centrum pro regionální rozvoj ČR** (regiony, regionální rozvojové agentury, regionální informační servis (RIS, <http://www.risy.cz>), Integrovaný regionální informační systém (IRIS), příhraniční regiony, euroregiony)

<http://epp.eurostat.ec.europa.eu/> **Eurostat** (statistiky EU. Eurostat Yearbook, strukturální indikátory).

<http://www.euroskop.cz/> **Úřad vlády ČR** (oficiální zdroj informací o členství ČR v EU, euroinfocentra (EIC), průzkumy veřejného mínění, analýzy)

<http://portal.uur.cz/> **Portál územního plánování** (dokumenty, obecné a oborové informace o území, územně identifikační registr, mapové podklady, orgány a instituce územního plánování, teorie, publikace, slovník)

<http://www.isu.cz/> **Informační portál o území (ISÚ)** (program obnovy venkova, www. stránky obcí ČR, programy regionálního rozvoje)

<http://portal.gov.cz/> **Portál veřejné správy** (ministerstva, EU, nahlížení do katastru nemovitostí, mapové služby)

<http://portal.mpsv.cz/> **Integrovaný portál MPSV** (brána do internetu, volná místa, měsíční a čtvrtletní statistiky nezaměstnanosti, insolventní firmy, nezaměstnanost z územního hlediska za okresy, zaměstnávání cizinců, regionální statistika ceny práce, uchazeči o zaměstnání a jejich podrobná struktura)

<http://www.demografie.info/> **Demografický informační portál** (analýzy, informace o úmrtnosti, nemocnosti, porodnosti, potratovosti, sňatečnosti, rozvodovosti, migraci, demografický slovník, mapy krajů a okresů, projekce ČR a OSN)

<http://www.uzis.cz/> **Ústav zdravotnických informací a statistiky MZ** (registr zdravotnických zařízení, údaje o zdraví – dynamické tabulky, Výběrové šetření o zdravotním stavu populace České republiky 1999 a 2002, Světové šetření o zdraví v České republice)

<http://www.mzcr.cz/> **Ministerstvo zdravotnictví** (informace pro širokou a odbornou veřejnost, Evropská unie)

<http://www.czso.cz/> **Český statistický úřad** [Ediční plán publikací (Regionální informace, Práce, sociální statistiky+práce a mzdy), Obyvatelstvo, volby Kraje (Krajské reprezentace ČSÚ), Registr ekonomických subjektů, Obyvatelstvo, Analýzy, komentáře, Časové řady, Regionální údaje, Statistická ročenka ČR, SLDB (Výsledky, Vyhledávání, Libovolné – všechny tabulky z posledního sčítání), Cizinci v ČR, Mezinárodní srovnání]

<http://www.epusa.cz/> **Elektronický portál územních samospráv** (okresy ČR, obce III. a II. stupně, správní obvody a úřady)

<http://mesta.obce.cz/> **Města a obce online** (mikroregiony jako svazky obcí)

<http://www.smocr.cz/> **Svaz měst a obcí České republiky** (dotace a granty, obce a EU)

<http://www.uur.cz/> **Ústav územního rozvoje Brno** (pod MMR, evidence územně plánovací einnosti obcí, krajů, mapový server, ročenka 2001, 2003 – aktuální stav urbanistických studií a územně plánovací dokumentace obcí a velkých územních celků v ČR, územní plánování, regionální politika (mikroregiony 2001-2004), bytová politika, cestovní ruch, obnova venkova, regenerace sídel, regenerace panových sídlišť, památková péče)

<http://www.mmr.cz/> **Ministerstvo pro místní rozvoj** (dokumenty a publikace MMR, regionální politika, regionální programy MMR, statistiky cestovního ruchu, statistiky bydlení, stavební spoření, hypoteční úvěry)

<http://www.mvcr.cz/> **Ministerstvo vnitra** (časopisy Veřejná správa, Správní právo, Moderní obec, azyl a migrace, integrace cizinců, dokumenty a koncepce, statistiky kriminality, nehodovosti, Centrální registr obyvatelstva)

<http://www.env.cz/> **Ministerstvo životního prostředí** (Natura 2000, Ekofilm, Indikátory životního prostředí, Mapové aplikace, Voda, Půda, Ovzduší, Odpady, Nevládní organizace a ekologická uskupení,)

<http://www.mdcr.cz/> **Ministerstvo dopravy** (Statistika dopravy, Statistika MHD, Jízdní řády)

<http://www.mkcr.cz/> **Ministerstvo kultury** (Statistika kultury, Výkaz o památkových objektech s kulturním využitím)

<http://www.mpo.cz/> **Ministerstvo průmyslu a obchodu** (Registr živnostenského podnikání,

<http://www.vojensky-zemepisny-ustav.org> **Vojenský zeměpisný ústav** v Praze(neoficiální stránky, odkaz geografie = Geografický server, na něm matematická, socioekonomická, regionální, fyzická, politická geografie, mapy ČR, mezinárodních integrací, slepé mapy, tématické mapy)

<http://www.czechtourism.cz/> **Česká centrála cestovního ruchu – CzechTourism** (statistiky, mezinárodní cestovní ruch, příjezdový cestovní ruch, výzkumy Czech Tourism, návštěvnost památek, domácí cestovní ruch, turistické regiony, profil návštěvníků v regionech, venkovská turistika)

<http://www.tourism.cz/> **Asociace turistických regionů České republiky (ATUR ČR)** - usiluje o dobré jméno a kvalitní rozvoj služeb cestovního ruchu v přirozených turistických regionech a oblastech a o jejich připravenost přijímat domácí a zahraniční turisty. Úkolem je hájit, prosazovat a zastupovat společné zájmy svých členů souvisejících s rozvojovými aktivitami turistických regionů, marketingem, tvorbou produktů cestovního ruchu, investicemi cestovního ruchu, dopravou, rozvojem služeb a vzdělávání lidských zdrojů. Aktivitu prosazovat u orgánů státní správy, územních aj. orgánů a organizací.

<http://portal.justice.cz/> **Oficiální server českého soudnictví** (Ministerstvo spravedlnosti – statistiky kriminálních trestných činů, Institut pro kriminologii a sociální prevenci, Obchodní rejstřík)

<http://www.natura2000.cz/> **Projekt Natura 2000** (evropsky významné lokality v ČR, ptačí oblasti)

<http://www.soc.cas.cz/> **Sociologický ústav AV** (odkazy na publikace a realizované sociologické výzkumy, SDA – Sociologický datový archiv, Socioweb, CVVM – centrum pro výzkum veřejného mínění)

<http://www.socioklub.cz> **Sdružení pro podporu rozvoje teorie a praxe sociální politiky**

<http://uzemi.oblibena.cz/> **Územně identifikační registr MV** (adresy, mapy, PSČ)

<http://katastr.cuzk.cz/> **Informační systém katastru nemovitostí ČR Českého úřadu kartografického a zeměměřičského** (dálkový přístup, přístup pro registrovaného uživatele)

<http://www.cnb.cz/cz/index.html> **Česká národní banka** (statistiky ČNB)

<http://www.kr-urady.cz/> **Asociace krajů ČR** (prezentace krajů, <http://www.municipal.cz/>)

<http://www.natur.cuni.cz/> **Přírodovědecká fakulta UK** (Geografie, Katedra sociální geografie a regionálního rozvoje, Geografie na internetu, Osobní stránky s geografickou tematikou)

<http://www.mikroregion.net/> **Portál mikroregionů**

<http://www.obecni-urad.net/> **Portál obecních úřadů**

<http://www.vupsv.cz/> **Výzkumný ústav práce a sociálních věcí**

<http://www.ikaros.cz/> **Ikaros** (časopis pro informační společnost)

<http://nros.cz/> **Nadace rozvoje občanské společnosti**

<http://ivris.fss.muni.cz/> **Institut pro výzkum reprodukce a integrace společnosti**

<http://isea-cz.org/> **Institut pro sociální a ekonomické analýzy**

<http://ceses.cuni.cz/> **Centrum pro sociální a ekonomické strategie** (publikace vize pro ČR)

<http://www.migraceonline.cz/> **Portál zahraniční migrace v ČR** (legislativa, statistiky)

<http://topregion.cz/> **MMR** (rozvoj lidských zdrojů, zaměstnatelnost)

<http://europa.eu.int/> **Portál EU**

<http://www.naseevropa.cz> **Český institut pro integraci EU**

<http://www.integrace.cz/> **Institut pro evropskou politiku Europeum**

<http://www.nvf.cz/> **Národní vzdělávací fond**

<http://www.kiseb.cz/> **Institut regionálních informací**

<http://oldmaps.geolab.cz/> **Kartografická laboratoř UJEP Ústí nad Labem** (staré mapy českých zemí z I. a II. vojenského mapování)

<http://www.stezka.cz/> **Naučné stezky v Česku**

2 MODUL 2 - OBECNÉ ZPŮSOBY UKLÁDÁNÍ DAT

S rozvojem výpočetní techniky vyvstal problém, jak ukládat pořízená data (programy, výsledky, vstupní údaje). V začátcích se používaly děrné štítky a děrné pásky. Obrovskou nevýhodou této metody byla nízká kapacita. Proto se postupně přes magnetickou pásku a disketu vyvinuly dnešní moderní média.

Obrovský nárůst ukládaných dat rovněž způsobil větší nároky na přenosové rychlosti, způsob ukládání dat a jejich zálohování.

2.1 Strategie ukládání dat

Strategie ukládání dat je nalezení souboru pravidel, doporučení, procesů a nástrojů pro ukládání velkého objemu dat. Zaměření analýzy je zejména pohled na rozvoj dat v následujících letech a uzpůsobení strategie ukládání dat budoucím nárokům uživatelů. Uplatnění strategie ukládání dat optimalizuje náklady na uložení dat a umožňuje efektivně řídit rozvoj datové základny a implementaci nových procesů (např. certifikace, nová legislativa, bezpečnost apod.) Objem dat se zvyšuje. V závislosti na růstu dat rostou i náklady na datovou infrastrukturu.

V souvislosti s rozvojem moderních technologií již není jednoduché rozhodnout jakým způsobem vybudovat či rozvíjet systém ukládání dat. K volbě správné strategie je nutné znát nejen současný stav technologií, procesů, a vazeb ohledně ukládání dat v organizaci, ale hlavně cíle a odhady budoucího vývoje v následujících letech. Hledat parametry budoucího rozvoje znamená pracovat s komplexními informacemi o budoucnosti – tedy informacemi od uživatelů IS, části strategických plánů rozvoje, technologickými parametry, legislativou, procesními schémata atd.

Uzpůsobením strategie ukládání dat současným a budoucím požadavkům lze dosáhnout výrazných úspor efektivním pořizováním vhodných technologií a služeb, nastavením optimalizačních procesů a správným managementem datové infrastruktury.

2.2 Datová média

Datové médium nebo také datový nosič, záznamové médium je paměťový nosič datových informací používající k záznamu dat určitý fyzikální princip.

Podle charakteru signálu, který je nosičem datového záznamu, existuje záznam digitální nebo analogový. Pokud je nosičem informace analogový signál, je potřeba použít vhodné modulace digitálních veličin.

Pro digitální záznam se digitální hodnota uloží v binární formě. Záznam dat na datovém médiu může být permanentní (trvalý), semipermanentní (přepisovatelný) nebo volatilní (nestálý, např. po vypnutí napájení se obsah ztratí). Mezi datová média můžeme zařadit všechny druhy a typy datových pamětí.

V praxi se pod pojmem datové médium často myslí přenosné výměnné datové médium. Tyto datové média jsou především určena k ukládání datových souborů a způsob uložení souborů na datovém médiu určuje typ použitého souborového systému.

2.2.1 Dělení datových médií

Podle principu čtení se datové nosiče dělí na:

Magnetická média, tzn. disketa, pevný disk, magnetooptický disk, magnetická páska (audiokazeta, videokazeta, DAT kazeta, LTO 1 až N)



Diskové pole

Optická média, tzn. CD, DVD, Blue-ray, HD DVD



Blu-ray disk

Elektronická média, tzn. flash paměť (Secure Digital, Multimedia Memory Card, Memory Stick, Flash card, xDcard, USB flash paměť)



Flash paměťová média

2.3 Komprese dat

Komprese dat (také komprimace dat) je speciální postup při ukládání nebo transportu dat. Úkolem komprese dat je zmenšit datový tok nebo zmenšit potřebu zdrojů při ukládání informací. Obecně se jedná o snahu zmenšit velikost datových souborů, což je výhodné např. pro jejich archivaci nebo při přenosu přes síť s omezenou rychlostí (snížení doby nutné pro přenos). Komprese může být nutná při omezené datové propustnosti, např. mobilní telefon komprimuje hovor pro přenos GSM sítí.

Zvláštními postupy – kódováním, které je dané zvoleným kompresním algoritmem – se ze souboru odstraňují nadbytečné informace, snižuje se entropie dat. Komprese dat lze rozdělit do dvou základních kategorií:

Ztrátová komprese – při kompresi jsou některé informace nenávratně ztraceny a nelze je zpět rekonstruovat. Používá se tam, kde je možné ztrátu některých informací tolerovat a kde nevýhoda určitého zkreslení je bohatě vyvážena velmi významným zmenšením souboru. Používá se pro kompresi zvuku a obrazu (video), při jejichž vnímání si člověk chybějících údajů nevšimne nebo si je dokáže domyslet (do určité míry).

Bezeztrátová komprese – obvykle není tak účinná jako ztrátová komprese dat. Velkou výhodou je, že komprimovaný soubor lze opačným postupem rekonstruovat do původní podoby. To je nutná podmínka při přenášení počítačových dat, kde by ztráta i jediného znaku mohla znamenat nenávratné poškození souboru.

2.3.1 Kompresní poměr

Kompresní poměr je podíl velikosti původních dat ku velikosti komprimovaných dat. Například při kompresi 10MB souboru do 2MB je poměr $10/2 = 5$ (tj. 5 : 1 – pět ku jedné, pětkrát zmenšeno). Kompresní poměr je ovlivněn volbou kompresního algoritmu i typem komprimovaných dat. Úspora místa je vyjádřena jako 1 – opačný poměr, v našem příkladě $1 - 2/10 = 0,8$ (tj. 80% úspora).

2.4 Databáze

Databáze slouží k uspořádávání uložených informací. Hlavní výhodou databázově uložených dat je možnost účinnější práce s uloženými daty.

Z hlediska způsobu ukládání dat a vazeb mezi nimi můžeme rozdělit databáze do základních typů:

1. Relační databáze
2. Hierarchické databáze
3. Síťová databáze
4. Objektové databáze
5. Objektově relační databáze

3 MODUL 3 - MODERNÍ TRENDY VE VYHLEDÁVÁNÍ NA INTERNETU

Vznik internetu a jeho rozvoj měl dalekosáhlé důsledky s nakládáním s informacemi. V mnohých ohledech došlo k velmi dramatickým změnám v každodenním životě nás všech. Hlavním aspektem těchto změn je dostupnost informací a jejich „aktualizovanost“. Dnes již nikoho nepřekvapí, že je možno sledovat aktuální pozici balíku doručovaného zásilkovou službou nebo poštou, že lze koupit či zarezervovat lísky do kina s přesným výběrem místa. Moderní operační systémy počítačů dokonce mají integrované například postranní lišty s aktuálním počasím ve zvolených místech.

Významným pokrokem bylo dramatické zvýšení kapacit záznamových zařízení. Srovnáme-li například kapacitu diskety s dnes běžnými flash paměťmi, jedná se o cca 1 000 násobný nárůst kapacity. Ještě větší kapacitu mají pevné disky. Do komerčního prodeje se dostávají disky s kapacitou 1TB, přičemž ještě před deseti lety byla dostupná kapacita 40 MB, což je cca 2 500 krát více.

V průběhu roku 2006 bylo na celém světě vytvořeno přibližně 160 exabajtů digitálních dat. Tyto data tvoří fotografie, videa, e-maily, telefonní hovory a další digitální obsah, se kterým se denně setkáváme. Každý takový digitální soubor pak je v průměru třikrát zreplikován.

Pokud bychom chtěli fyzicky vyjádřit takové množství zaznamenaných informací, bylo by to něco jako 12 souvislých řad knih mezi Sluncem a Zemí, nebo třímilionkrát více knih, než kdy bylo vydáno. Pro 161 exabajtů by bylo třeba 2 miliard iPodů s nejvyšší kapacitou.

S tímto nárůstem dat se stále více do popředí dostává problém nejen s ukládáním a archivací dat, ale hlavně s tříděním, indexováním a vyhledáváním informací. Hlavním aspektem internetu je jeho decentralizace, která rovněž vede k neřízenému vytváření obsahu. Relevantní obsah, který hledáme, může být na kterémkoliv místě, jak na oficiálních stránkách města, v blogu neznámého autora či v elektronické encyklopedii na druhém konci zeměkoule.

Smyslem následujícího textu je popsat základy vyhledávání na internetu s pokročilejšími technikami. Rovněž se v něm krátce zmíníme o tom, jak vyhledávače fungují a jak připravit obsah, který publikujeme na internet tak, aby byl lehce vyhledatelný.

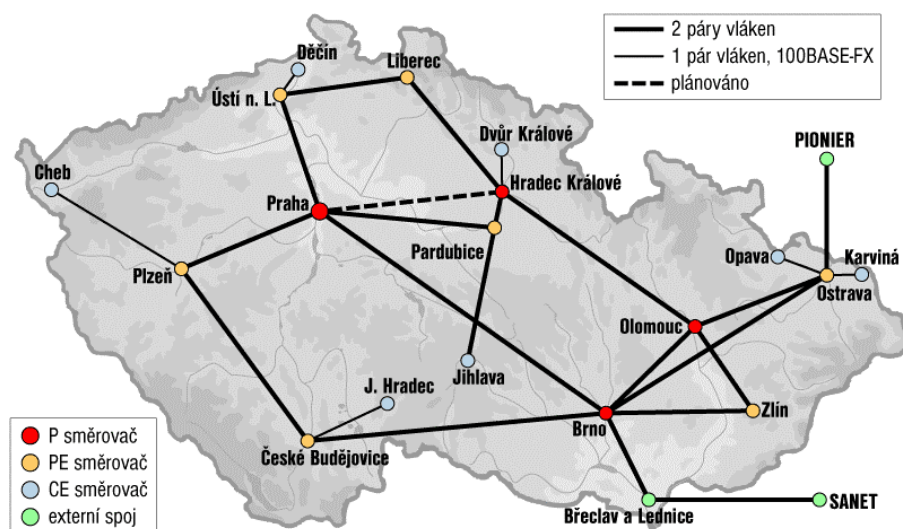
3.1 Internet a jeho služby

Internet je celosvětová počítačová „supersít“, která spojuje jednotlivé menší sítě pomocí sady protokolů IP.

Název pochází z anglického slova network (sít'), podle něhož tradičně názvy amerických počítačových sítí končily „-net“, a mezinárodní (původně latinské) předpony inter- (mezi), vyjadřující, že internet propojil a vstřebal různé starší, dílčí, specializované, proprietární nebo lokální sítě. Internet slouží k přenášení informací a poskytování mnoha služeb, jako jsou elektronická pošta, chat, www stránky, sdílení souborů, on-line hraní her, vyhledávání, katalog a další.

3.1.1 Historie internetu

V šedesátých letech se americká armáda snažila najít způsob, jak zajistit, aby armádní počítače rozmístěné po celém území USA mohly spolu bez problému komunikovat, a to i v případě, že část této sítě bude vyřazena z provozu. Pracovníci RAND Corporation přišli s unikátním řešením - vybudování sítě bez centrálního uzlu. Pokud bude některá linka zničena, informace bude ihned vedena k příjemci jinou trasou. "Are you receiving this?" - první věta, která byla v srpnu 1969 poslána z University of California v Los Angeles po síti složené ze čtyř uzlů: UCLA, Stanford Research Institute, UC Santa Barbara a University of Utah v Salt Lake City. Tak vznikl arpanet.



Decentralizovaná páteří síť CESNET

Postupně se k internetu připojovaly další instituce, především univerzity. V této době byl internet čistě nekomerční záležitostí. Na jeho vybudování přispívala americká armáda a různé vládní agentury. Pro komerční účely nebyl ani použitelný.

V roce 1989 vytvořil Tim Berners-Lee nový způsob sdílení dokumentů (původně pro vnitřní potřebu laboratoří CERN, kde pracoval) - hypertextové dokumenty. Texty, které obsahují odkazy na další dokumenty, které mohou být umístěny na jiném počítači, třeba na druhém konci světa. Díky jednoduchému a intuitivnímu ovládání se tento způsob komunikace rozšířil i za brány CERNu a dnes jej známe pod jménem World Wide Web. Zanedlouho byly k dokumentům připojeny i obrázky. Vzhled dokumentů byl přirozenější a umožnil ještě lepší komunikaci. Právě existence www spolu s masovým rozšířením osobních počítačů přilákala k internetu miliony nových uživatelů, a tím začal být internet zajímavý i pro podnikatelský sektor. Komerční provoz na internetu se datuje od roku 1992, kdy National Science Foundation, která do této doby spravovala páteří síť internetu, umožnila připojení i komerčním subjektům.

3.1.2 Základní služby (protokoly)

Jednotlivé služby internetu používají svoje protokoly. Díky jejich diferenciaci je možno lepší rozdělení zátěže, priorit a zabezpečení serverů i klientů.

SMTP - Simple Mail Transfer Protocol, elektronická pošta, e-mail

FTP - File Transfer Protocol, přenos vzdálených souborů

Telnet - Virtuální terminál, vzdálený přístup

NFS - Network File System, sdílení vzdálených souborů

DHCP - Dynamic Host Configuration Protocol, dynamická konfigurace síťové stanice

SNMP - Simple Network Management Protocol, jednoduchý protokol pro správu sítě

HTTP - Hypertext Transfer Protocol, World Wide Web

DNS - Domain Name System, překlad doménových jmen

Z hlediska internetu a běžného provozu je nejpodstatnější protokol HTTP a případně jeho šifrovaná verze HTTPS. Jejich prostřednictvím dochází k sdílení obsahu. Zcela mylné je povědomí mnoha uživatelů, že internet je vlastně WWW, zatímco email, ftp a další služby internet nejsou. Tento omyl vzniká hlavně z toho důvodu, že právě na webu je dostupná většina obsahu. Dnešní prohlížeče a webové aplikace v sobě integrují ale i mnohé jiné služby a protokoly, takže běžný uživatel ani nemusí poznat, že nevyužívá protokol HTTP ale např. i FTP.

3.1.2.1 Protokol http

HTTP (Hyper Text Transfer Protocol) je internetový protokol určený původně pro výměnu hypertextových dokumentů ve formátu HTML. Tento protokol je spolu s elektronickou poštou tím nejvíce používaným a zasloužil se o obrovský rozvoj internetu v posledních letech.



V současné době je používán i pro přenos dalších informací. Pomocí rozšíření MIME umí přenášet jakýkoli soubor (podobně jako e-mail), používá se společně s formátem XML pro tzv. webové služby (spouštění vzdálených aplikací) a pomocí aplikačních bran zpřístupňuje i další protokoly, jako je např. FTP nebo SMTP.

K protokolu HTTP existuje také jeho zabezpečená verze HTTPS, která umožňuje přenášet data šifrovat a tím chránit před odposlechem či jiným narušením.

3.1.2.1.1 Činnost protokolu

Protokol funguje způsobem dotaz-odpověď. Uživatel (pomocí programu, obvykle internetového prohlížeče) pošle serveru dotaz ve formě čistého textu, obsahujícího označení požadovaného dokumentu, informace o schopnostech prohlížeče apod. Server poté odpoví pomocí několika řádků textu popisujících výsledek dotazu (zda se dokument podařilo najít, jakého typu dokument je atd.), za kterými následují data samotného požadovaného dokumentu.

Pokud uživatel bude mít po chvíli další dotaz na stejný server (např. proto, že uživatel v dokumentu kliknul na hypertextový odkaz), bude se jednat o další, nezávislý dotaz a odpověď. Z hlediska serveru nelze poznat, jestli tento druhý dotaz jakkoli souvisí s

předchozím. Kvůli této vlastnosti se protokolu HTTP říká bezstavový protokol – protokol neumí uchovávat stav komunikace, dotazy spolu nemají souvislost. Tato vlastnost je nepříjemná pro implementaci složitějších procesů přes HTTP (např. internetový obchod potřebuje uchovávat informaci o identitě zákazníka, o obsahu jeho „nákupního košíku“ apod.). K tomuto účelu byl protokol HTTP rozšířen o tzv. HTTP cookies, které umožňují serveru uchovávat si informace o stavu spojení na počítači uživatele.

WWW (World Wide Web) – informační systém pro práci s hypertextovými dokumenty, ve kterých jsou odkazy na internetovské zdroje uváděny pomocí adresy. Je to nejrozšířenější služba v současném internetu. Jistým nedostatkem je archivace zveřejněných dokumentů – lze je libovolně měnit a nelze tedy zaručit, že se někdo např. nebude pokoušet falšovat historii. Rovněž se není možné spolehnout, že dříve zveřejněný dokument bude na internetu stále k dispozici a k nalezení.

3.1.3 Druhy a formáty internetového obsahu

Internet svou technickou podstatou umožňuje distribuovat data v elektronické podobě. Hlavním médiem (a vlastně původním) je textový obsah. S nástupem multimédií do běžného života a digitalizací většiny oblastí života došlo k rozšíření i na obrazovou a zvukovou složku.

3.1.3.1 Text

Základním médiem obsahu internetu byly textové informace. Jak již bylo zmíněno hlavním důvodem vzniku WWW byla potřeba zveřejňovat a sdílet textové informace.

3.1.3.1.1 Prostý text

Nejjednodušším typem textového formátu je prostý text. Neobsahuje žádné formátovací značky mimo konce řádků a několik dalších základních odstavcových značek.

3.1.3.1.2 Rich Text Format

Je to Microsoftem vyvinutý, na platformě nezávislý formát souboru pro uložení textu, který obsahuje co největší množinu formátovacích příkazů. Vznikl v roce 1987. Pomocí tohoto formátu je obecně možné vyměňovat dokumenty mezi nejrozličnějšími programy pro zpracování textu se zachováním vzhledu a formátu. Na rozdíl od většiny vlastních formátů souborů textových editorů je RTF čitelný i v prosté textové podobě, tedy jeho obsah vypadá jako zvláštní text ASCII, nikoliv jako změř nesmyslných znaků.

3.1.3.1.3 Hypertext

Kvůli potřebě provázat jednotlivé texty vznikly různé hypertextové formáty. Jedná se o strukturovaný elektronický text, obsahující odkazy na jiné texty, obrázky, zvuky, animace, video. Používá se na internetu, ale i lokálně (encyklopedie, nápovědy atd.) Dnes nejznámější a nejpoužívanější je HTML a XML.

3.1.3.1.4 PDF

PDF (zkratka anglického názvu Portable Document Format – Formát pro přenositelné dokumenty) je souborový formát vyvinutý firmou Adobe pro ukládání dokumentů nezávisle na softwaru i hardwaru, na kterém byly pořízeny. Soubor typu PDF může obsahovat text i obrázky, přičemž tento formát zajišťuje, že se libovolný dokument na všech zařízeních zobrazí stejně. Pro tento formát existují volně dostupné prohlížeče pro mnoho platform, nejznámějším je oficiální prohlížeč mateřské firmy Adobe – Adobe Reader.

3.1.3.1.5 Ostatní formáty

Z ostatních formátů pro textové dokumenty je možno uvést například proprietální formát MS Office, nově budovaný formát Open Document Standard. Jejich použití pro publikování informací na internetu je však značně problematické, i když dnešní vyhledávače již dokáží zaindexovat i tyto dokumenty.

3.1.3.2 Zvuk

Audio informace patří k dalšímu formátu, který je možno na internetu poskytovat. Dnes nejběžnější formáty k distribuci internetem jsou MP3, WMA a OGG.

3.1.3.3 Grafika

Grafika dnes doplňuje většinu internetového obsahu. Používají se většinou bitmapové formáty JPG, GIF a PNG.

3.1.3.4 Video

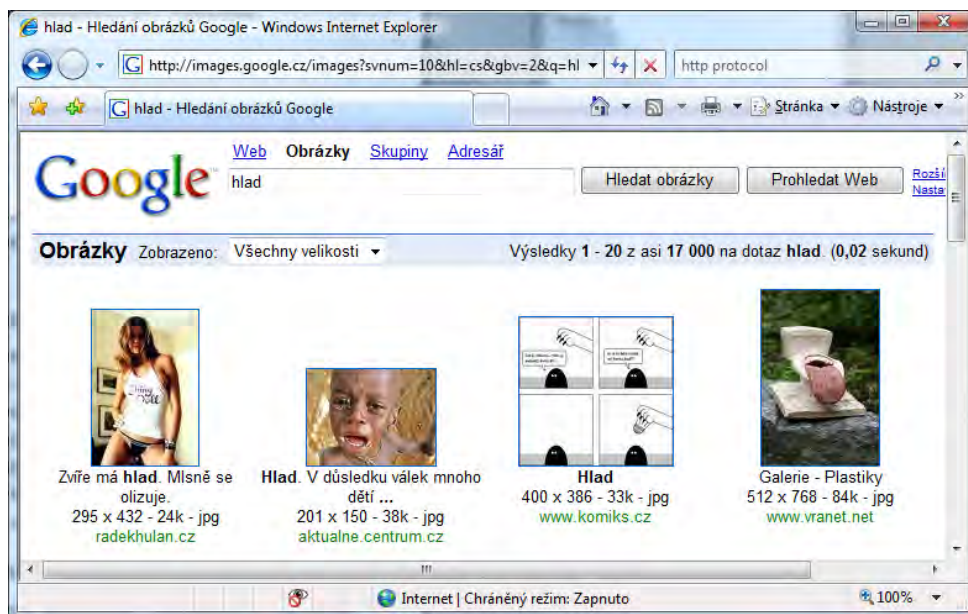
Spojení obrazu a zvuku je nejnáročnější na datové kapacity internetu a proto k jeho rozvoji dochází až v poslední době. Formátů je celá řada, jmenujme např. FLV, MPEG nebo MOV.

3.1.3.5 Indexování textových dokumentů

Indexování textových dokumentů, a jejich pozdější vyhledávání je v podstatě nejjednodušší a jediné doposud uspokojivě vyřešené. Existuje mnoho nástrojů na analýzu textu, která dokáže automaticky rozpoznat jazyk, obsah i kvalitu textového dokumentu. Těmž pomocí jistých pravidel dokáže určit jeho věrohodnost a originalitu a rozhodnout o jeho informační hodnotě. Pro zaindexování textů na internetu je ideální některý hypertextový formát, který umožňuje text opatřit značkami, které napomáhají správnému zaindexování.

3.1.3.6 Indexování ostatních médií

Indexování ostatních médií je značně problematické. Vývoj softwaru na rozpoznávání řeči je zatím spíše ve vývoji a i tak jeho výsledky ještě nejsou vhodné pro nasazení na vyhledávání. U obrazového materiálu je situace ještě horší. V obou případech tedy slouží k indexování textové informace k médiu (například hlavičky souborů). Správné zaindexování multimediálního obsahu je v současné době poměrně problematické.



Příklad hledání obrázku na klíčové slovo HLAD. První obrázek je z webu, který pojednává o sexuálním hladu majitele webu, který ovšem nemá s obrázkem nic společného.

3.1.4 Dostupnost internetového obsahu

Obsah zveřejněný na internetu nemusí být zaindexovaný, tedy dostupný pro vyhledávače, a tedy zaindexovaný z mnoha důvodů. Při vyhledávání tedy musíme vzít v úvahu, že hledaný obsah může existovat v takzvaném Hlubokém webu.

Hlavní důvody neindexování obsahu stránek jsou tyto:

3.1.4.1 Obsah se nachází v Hlubokém webu

Hluboký, neviditelný web je označení pro stránky a zdroje na internetu, které nenajdeme běžnými vyhledávacími nástroji (např. Google). Neviditelný web tvoří z největší části informace skryté ve specializovaných databázích a stránkách. Může jít například o vědecké zdroje univerzit, firemní stránky a databáze. Neviditelný web je údajně až 500krát obsáhlejší než web povrchový - prohledatelný běžnými vyhledávacími nástroji.

3.1.4.2 Stránky malé, špatně přístupné

Mnohdy se informace nacházejí na špatných, malých, nebo hluboko vnořených stránkách. Může se jednat například o informace lokálního významu, například obecní stránky, na které nevede žádný zpětný odkaz.

3.1.4.3 Privátní nebo zaheslované stránky

Některé WWW stránky jsou záměrně nepřístupné. Může se jednat například o firemní informační stránky, ale i o komerční projekty s placeným přístupem.

3.1.4.4 Databázově orientované stránky

Jedná se například o katalogy, které jsou dostupné po zadávání klíčových slov. Roboti, kteří indexují www stránky, ovšem nevyplňují formuláře, a proto jim může zůstat tato část webu nedostupná.

Obecně platí, že komerční projekty, jako například e-shopy, katalogy, zpravodajské servery apod., bývají pro vyhledávače přístupnější, respektive většinou lépe optimalizované. Naopak informačně mnohdy „hodnotnější“ stránky, jako například vědecké práce, odborné články a lokální informace bývají více nedostupné. Kromně komerčního důvodu je příčinou často neznalost autora nebo případně slabá motivace pro zpřístupnění obsahu. Příkladem může být výsledek vyhledávání hesla „Fotoaparát Nikon“. V první řadě e-shopy s fototechnikou a teprve po nich se nám zobrazí informace o výrobku z recenzí. Překvapivě oficiální stránky značky na předních pozicích chybí.

3.2 Vyhledávače

Internetový vyhledávač je stránka, určená k vyhledávání hlavně dalších stránek, ale také souborů (např. obrázků, nebo audia) na internetu.

Program internetového vyhledávače po zadání hledaného výrazu hledá ve své zaindexované databázi uživatelem zvolený termín. Protože všechny vyhledávače spolu soutěží o návštěvnost, nabízejí různé doplňkové služby, jako je například podrobnější vyhledávání, zpravodajství nebo e-mailová schránka zdarma.

3.2.1 Vyhledávač/katalog/meta vyhledávač/Open Directory Project

Z hlediska způsobu funkce vyhledávače je můžeme rozdělit na vyhledávače a katalogy. Katalog na rozdíl od vyhledávače pracuje s daty vloženými do příslušné kategorie. Umístění odkazu většinou nezávisí na obsahu stránky ale na zařazení do příslušné kategorie a správném popisku. Příkladem může být například katalog.seznam.cz. Tento katalog trpí bohužel obecným problémem neaktuálnosti a nerelevantnosti dat. Rovněž není schopen postihnout rozsáhlé a dynamické weby, jako například zpravodajské servery apod. Naopak vyhledávač pracující automaticky se snadněji stane obětí podvodu v podobě nepovolených SEO technik.

Jistým specifickým katalogem je Open Directory Project. Patří mezi největší katalogové servery na internetu. Vznikl jako vedlejší produkt při vývoji prohlížeče Mozilla a právě Mozilla je jeho maskotem. Proto se mu také někdy říká dmoz - Directory of MOZilla.

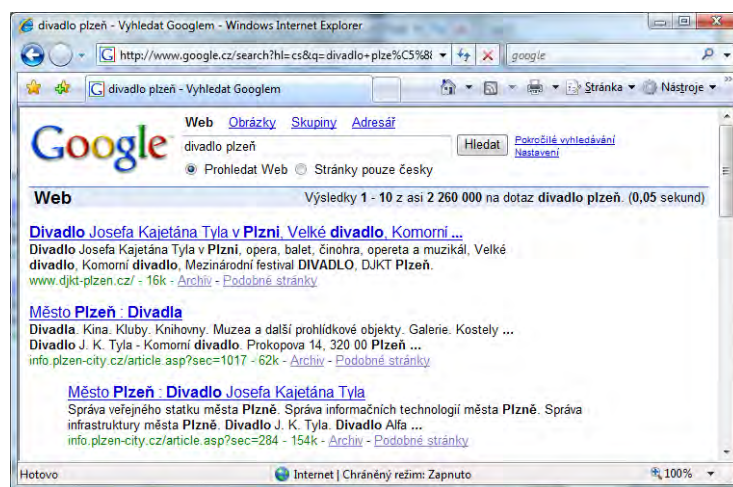
V současné době je význam katalogů na ústupu. Jeho hlavní význam je pro subjekty, které nemají vlastní internetové stránky (například řemeslníci nebo firmy s lokálním významem). Pro ostatní slouží spíše ke získávání zpětných odkazů.

Jistou alternativou k výše uvedeným jsou Metavyhledávače. Metavyhledávače jsou servery, které nemají svůj vlastní vyhledávací stroj, ale používají a sdružují při zadání dotazu vícero vyhledávačů z jiných serverů (např. Google, Altavista atd.). Výhody Metavyhledávačů jsou jednak ve velké úplnosti prohledávaných informací a také v možnosti integrovat libovolné vyhledávací služby do jediné. To může být výhodné pro firemní vyhledávací brány. Metavyhledávač nemusí také stahovat webové stránky k indexování, takže je, co se týče provozních nákladů, přívětivější.

3.2.2 Jak funguje vyhledávač

Většina internetových vyhledávačů pracuje ve třech krocích - prohledání webu, indexování a vyhledávání.

Pro prohledání webu má internetový vyhledávač robota (tzv. bot), který prochází celý web doslova odkaz za odkazem a zaznamená si obsah navštívených stránek. Tento zaznamenaný soubor poté zanalyzuje a provede tzv. indexaci, čili stáhnutí celé stránky na vlastní server, a nachystá její začlenění do databáze vyhledávání. Z tohoto postupu plyne důvod, proč se změny webu neprojeví ve výsledcích vyhledávání okamžitě, ale až po určité době (až bot stránku znovu zaindexuje). Následně koncový uživatel zadá do vyhledávače klíčová slova a vyhledávač navrátí shodu klíčových slov se stránkami v databázi, většinou v přehledné formě jako je titulek stránky, její obsah a adresa.



Užitečnost vyhledávače tvoří z velké části relevance nalezených stránek k hledanému tématu. Proto většina vyhledávačů přiřadí stránce podle mnoha kritérií určité hodnocení (PageRank u Google, S-Rank u Seznamu), a z toho poté vychází při třídění výsledků vyhledávání. Stránku, na kterou vede pouze minimum odkazů, bude považovat za méně důležitou než stránku s tisíci zpětnými odkazy, a proto ji zařadí níže ve výsledcích vyhledávání.

Z hlediska techniky běží vyhledávače na velmi výkonných strojích schopných zpracovat miliony požadavků za sekundu. Google, jako světová jednička mezi vyhledávači, je provozován na více než 450 000 počítačích po celém světě. Jednotlivé počítače jsou většinou součástí lokálního clusteru - supervýkonného počítače. Každý server má přesně danou svojí úlohu (cache servery, databázové servery, web servery). Indexování celého internetu má také obrovské nároky na kapacitu disků.

3.2.2.1 Hodnocení stránek

V databázi internetového vyhledávače se pro běžné dotazy najde obrovské množství stránek, které zadání dotazu vyhovují. Úkolem vyhledávače je posoudit kvalitu a relevantnost všech těchto stránek a prezentovat je uživateli seřazené od těch nejdůležitějších, nej kvalitnějších.

Přesné metody a algoritmy, podle nichž vyhledávač stránky hodnotí, jsou obvykle přísně střeženým tajemstvím. Pozice stránky ve výsledcích vyhledávání závisí na tom, do jaké míry podle algoritmu vyhledávacího stroje splňuje očekávání vyhledávajícího uživatele. Na základě obecně známých informací o těchto algoritmech lze odvodit, že pozici stránky ve výsledcích vyhledávání může napomoci, pokud na stránku odkazuje hodně dalších stránek,

pokud se vyhledávaný výraz na stránce vyskytuje vícekrát, na důležitých místech (např. v titulku) apod. Rovněž je zohledňována kvalita textu (jazyková) a unikátnost.

3.2.2.2 Page Rank

PageRank (PR) používá Google (GPR) a mnoho jiných vyhledávačů. PR vyjadřuje věrohodnost, důležitost stránky a ovlivňuje řazení výsledků. Page Rank konkrétních stránek je možno zjistit pomocí nástroje Google Toolbar.

GPR se počítá podle toho, kolik stránek danou stránku odkazuje -- čím více, tím lepší PageRank. Odkazy ze stránek s vyšším PR mají pochopitelně větší váhu.

Český vyhledávač Jyxo provozuje i JyxoRank, Seznam.cz zase S-rank, které se počítají podle toho, z kolika různých domén druhé úrovně vedou na stránku odkazy.

3.2.2.3 Roboti

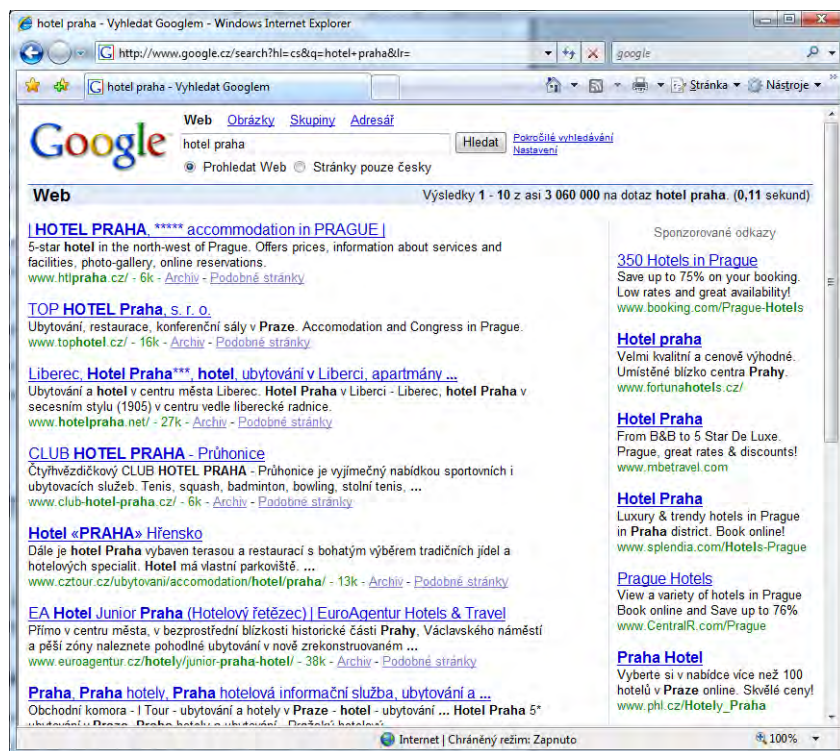
Programům, které procházejí web a schraňují pro vyhledávače informace, se říká roboti, boti nebo crawleři. Tento typ robotů prochází jednotlivé webové stránky, hledá na nich odkazy na nové stránky, indexuje obsah zpracovávaných stránek a umožňuje jejich následné prohledávání.

3.2.2.4 Obchodní model vyhledávače

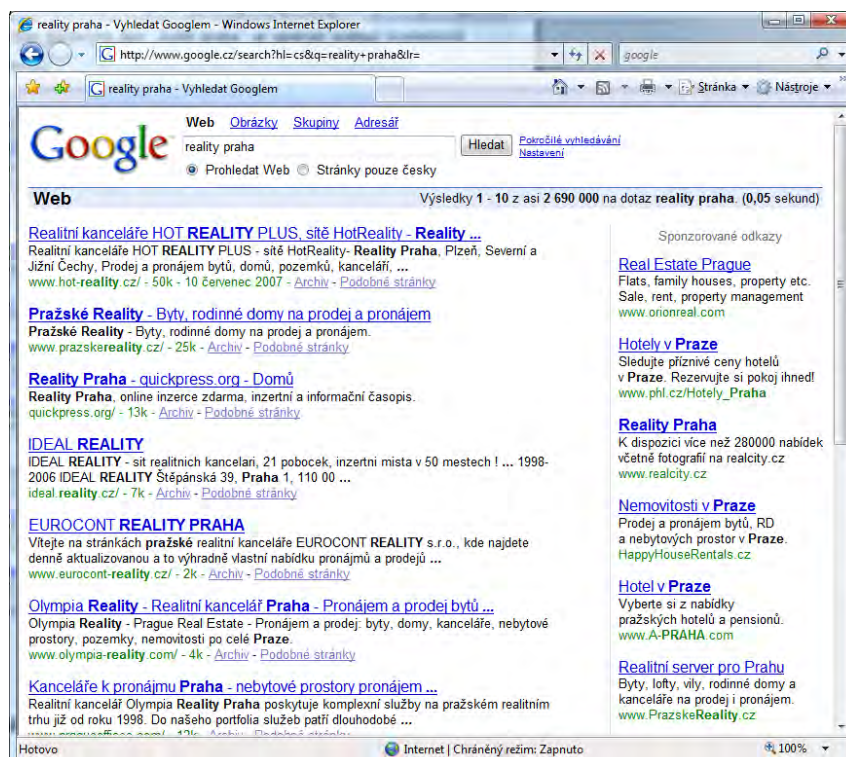
Přestože se zdá, že programování a provoz vyhledávače není příliš výnosné, opak je pravdou. Tento druh služby má určitá pravidla, která je třeba znát pro pochopení fungování této služby.

Vyhledávače nenabízejí služby pro uživatele za peníze, jinými slovy jsou zdarma. Nicméně provozovatel vyhledávače profituje z reklamy, např. takzvaných kontextových odkazů, sponzorovaných odkazů nebo bannerové reklamy. Forem této reklamy může být více, podstatné je ale to, že i tato reklama musí být relevantní k hledaným klíčovým slovům. Je však zřejmé, že reklamní plochu prodá pouze vyhledávač s dobrou návštěvností, kterou lze zabezpečit kvalitním vyhledávacím nástrojem.

Pokud vyhledávač bude špatně indexovat stránky ztratí návštěvníky a poté i inzerenty. Rovněž pokud sponzorované odkazy budou nerelevantní, potenciální zákazník nevyužije nabízené služby, a pro inzerenta nedojde ke kýžené konverzi nakoupené reklamy. Například při vyhledání klíčových slov „hotel Praha“ se správně zobrazí kontextová reklama na pražské hotely.



Na druhou stranu se podobné reklamní texty zobrazí i na dotaz „reality Praha“.



Je ovšem velmi pravděpodobné, že zákazník hledající nemovitost v Praze nehledá hotel, a i když náhodou na reklamu klikne, nevyužije hotelových služeb. Dobrý vyhledávač sleduje kvalitu reklamních textů a snaží se různými metodami o jejich optimalizaci.

3.2.2.5 Světové vyhledávače

Boj o prvenství mezi světovými vyhledávači jednoznačně vede Google. V některých zemích, například na Slovensku, není ani jiný významný lokální vyhledávač, kterým je u nás Seznam.cz.

Vyhledávač	Podíl
Google	47,7 %
Yahoo!	21,8 %
MSN	11,3 %
AOL	7,2 %
Ask Jeeves	2,6 %

3.2.2.5.1 Google

je největší světový internetový vyhledávač, probíhá na něm 47,7% vyhledávání. Americká společnost Google, Inc., která ho provozuje, sídlí v Mountain View v Silicon Valley jižně od San Franciska v Kalifornii. Vyhledávač každodenně obslouží přes 200 milionů dotazů. Kromě webových stránek umí prohledávat také obrázky, diskusní skupiny, zpravodajské servery a nabídky on-line prodeje. V červnu 2005 dokázal Google hledat v 8,05 miliardách stránek, 1,3 miliardách obrázků a více než miliardě zpráv z diskusních skupin. Většinu tohoto obsahu také archivuje. Google nabízí své rozhraní v mnoha jazycích, včetně češtiny.

Popularita Googlu je taková, že se dnes již používá i sloveso googlovat či googlit ve smyslu hledat na internetu. V roce 2006 se toto sloveso objevilo v Oxford English Dictionary.

3.2.2.6 České vyhledávače

Zatímco na mezinárodním poli jsou zajímavé většinou jen největší vyhledávače, v českých podmínkách je třeba se zmínit i o některých „oborových“ či speciálních variantách vyhledávání. Důvodem je především lokálnost nabízených služeb. Význam neglobálního vyhledávání se rozšiřuje totiž i na neelektronické formy služeb, například zboží, ubytování či jiné služby. Proto přichází ke slovu i zdánlivě méně použitelné katalogy.

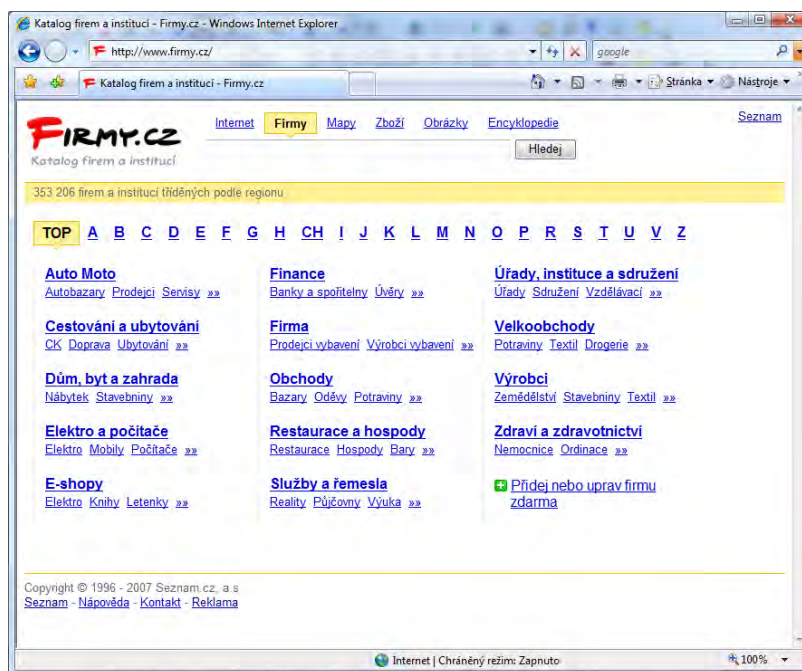
3.2.2.6.1 Fulltextové vyhledávače

Pro internet v naší zemi co se týče vyhledávačů, platí poněkud mimořádná situace. Zatímco jinde mají lokální firmy poskytující tyto služby menšinové zastoupení, u nás je dominantní vyhledávač Seznam.cz. Ten poskytuje jak fulltext tak katalog. Ve fulltextu za ním následují Centrum.cz se svým vyhledávačem Morfeo a Atlas.cz, který k fulltextovému vyhledávání používá služeb vyhledávače Jyxo.cz.

3.2.2.6.2 Katalogy

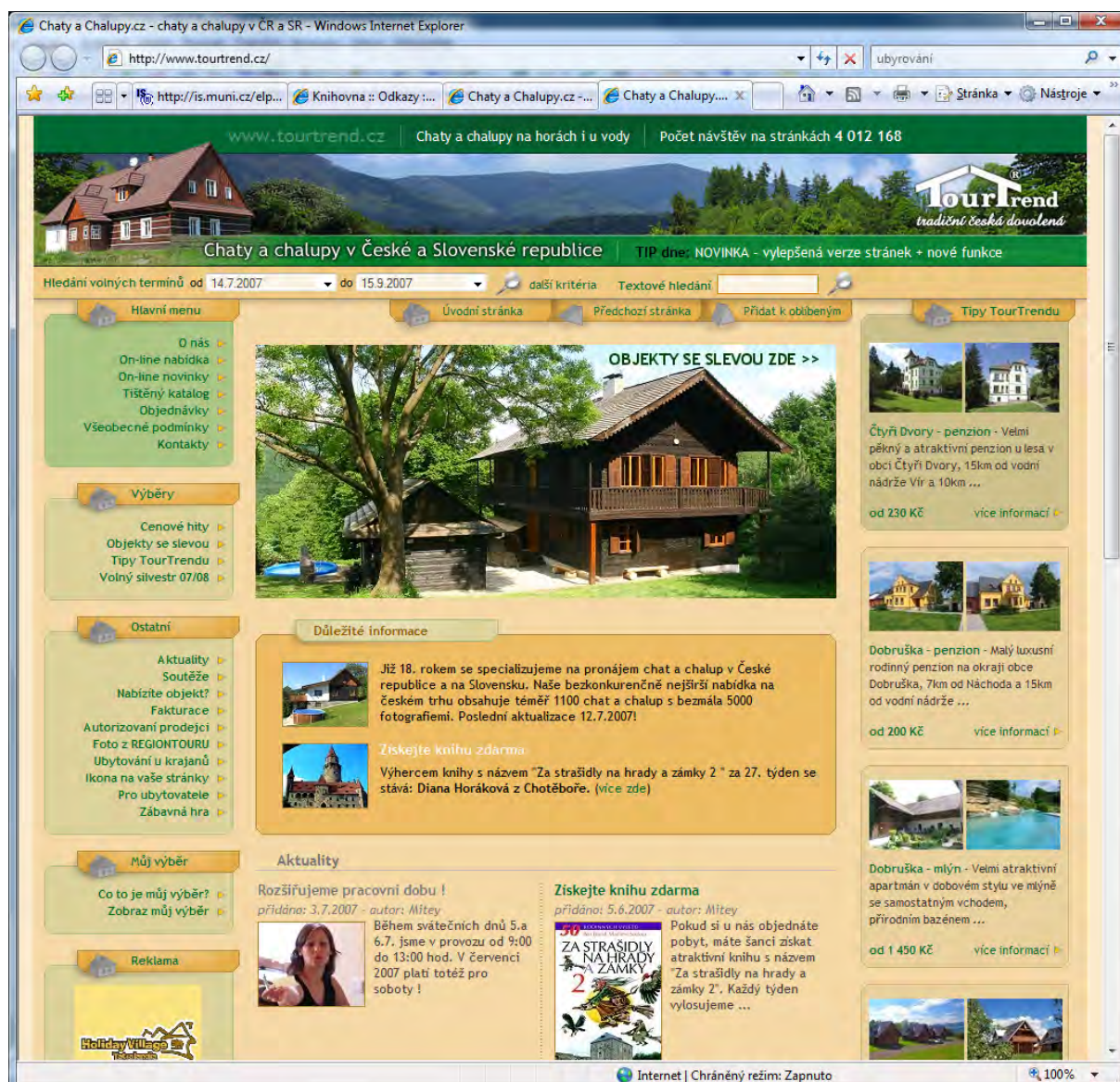
Opět nejznámějším katalogem je Seznam.cz. Výhodou katalogu je, jak již bylo zmíněno, že může obsahovat i informace od subjektů, nemajících vlastní obsah na internetu. Výhodou katalogů je rovněž možnost členit data lokálně, například podle okresů. Katalog bývá

rozdělen do sekcí, kterými je nutné procházet. Nevýhodou tohoto systému je, že vzhledem k množství hesel v katalogu je hledaná informace vložena poměrně hluboko a cesta k ní nemusí být správně určena.



Katalog seznamu – úvodní stránka

Z dalších významných katalogů můžeme zmínit například zlatestranky.cz, které jsou elektronickou verzí telefonního seznamu. Z oborových vyhledávačů můžeme zmínit například www.obchodnirejstrik.cz, který slouží k vyhledávání firem, www.cuzk.cz, který umožňuje nahlížení do katastru nemovitostí. Vyhledávání ubytování například slouží servery www.tourtrend.cz, www.pampeliska.cz/ubytovani/vyhledavac/, www.cz-hotel24.com a mnohé další.



Příklad vyhledávače ubytování

Katalogy nicméně mohou konkurovat fultextovým vyhledávačům jen částečně a to právě v jednotlivých oborech.

3.2.2.6.3 Speciální vyhledávače

Většina vyhledávačů je zaměřena na vyhledání textové informace. Některé vyhledávače mohou být zaměřeny oborově, některé vyhledávače poskytuje státní správa, jako například registr plátců DPH a podobně. Existují ale i jiné, než textové informace. Příkladem může být služba www.mapy.cz, mapy.atlas.cz apod., které dokáží vyhledávat na mapách. Některé mapové vyhledávače rovněž dokáží plánovat trasu. Dalším příkladem může být vyhledávání obrázků nebo hudby. Jistý specifický, byť Metavyhledávač je například služba vyhledávání zboží, která dokáže vyhledat nabídku zboží v jednotlivých obchodech a představit nejlevnější z nich.

Seznam Zboží - Seznam Zboží - Windows Internet Explorer

http://zbozi.seznam.cz/searchScreen?q=nikon%20d40x&order=cheapest&minPrice=&maxPrice= google

Seznam Zboží

Internet Firmy Mapy Zboží Obrázky Encyklopedie Seznam

nikon d40x Hledej

Vyhledávám "nikon d40x" Zobrazuji 61 - 70 z 725 nalezených

Seřadit dle:

☐ Nejlepší shody

☒ Od nejnižšího

☐ Od nejdražšího

Cenové rozpětí:

Do 1 000,- Kč | 1 000,- až 3 000,- Kč | 3 000,- až 5 000,- Kč | 5 000,- Kč a výše

Od: Kč Do: Kč OK



Fotoaparát digitální zrcadlovka Nikon D40X
19 371,- Kč s DPH - Elektro Cigler, Brno
 Fotoaparát D40x je lehký, snadno použitelný a dobře přenosný přístroj, se kterým je rovněž velmi zábavné fotografovat. Stačí jen nastavit expoziční režim AUTO, podívat se jasným a ostrým hledáčkem...



Nikon digitální zrcadlovka D40X Body
19 371,- Kč s DPH - Best buy
 Fotoaparát D40x ... D40x reaguje tak rychle, že umožní ...



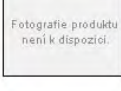
Nikon digitální zrcadlovka D40X Body
19 371,- Kč s DPH - Probit elektro
 Fotoaparát D40x ... D40x reaguje tak rychle, že umožní ...



Nikon DZ D40X
19 387,- Kč s DPH - České Počítače s.r.o.
 Fotoaparát D40x ... D40x reaguje tak rychle, že umožní ...



Nikon D40x Body
19 390,- Kč s DPH - Obchodní-dům.cz
 Digitální fotoaparát - Obchodní-dům.cz - bílé zboží, domácí spotřebiče, elektronika



Nikon digitální zrcadlovka D40X Body
19 402,- Kč s DPH - Shop.cz
 Fotoaparát D40x ... D40x reaguje tak rychle, že umožní ...

Reklama Sklik:

Nikon D40X cash back 2500
 Digitální zrcadlovka Nikon D40X na Vás čeká v našem obchodě.
www.vkfoto.cz

Fotoaparáty Nikon skladem
 Kompletní nabídka fotoaparátů Nikon za super ceny na prodejní i e-shop
www.antikfoto.cz

Nikon D40X
 Nakupte Nikon D40X v zavedeném obchodě za výhodnou cenu!
www.Megapixel.cz

Nikon
 Kvalitní digitální fotoaparáty mnoha značek. Kompakty i zrcadlovky
www.FotoSkoda.cz

Nikon d40x
 Kupte si Nikon d40x za 21 840 Kč nebo na splátky.
foto.x-store.cz/nikon

Nikon
 Špičková fototechnika Nikon záruka nízké ceny a pohodlný nákup.
www.FotoExpert.cz

Fotoaparáty Nikon
 Digitální fotoaparáty Nikon. Velký výběr, příznivé ceny.
www.Aaron.cz/Nikon

Akce - Nikon D40x

Hotovo Internet | Chráněný režim: Zapnuto 100%

Vyhledávání zboží na zbozi.seznam.cz

3.3 Pokročilé vyhledávání

K vyhledání informace může vést vždy několik cest. Hlavním předpokladem je správné stanovení metody vyhledávání a posléze použití příslušného vyhledávače. Ne vždy musí být nejlepší řešení použití fulltextového vyhledávače. Typickým případem může být hledání sklenáře v konkrétním městě. Použití „kanónu“ v podobě Google může vést k nepříjemnému překvapení v podobě nalezení spousty neaktualizovaných údajů na různých, dnes již třeba neexistujících stránkách. V tomto případě je lepší použít například již zmíněné zlaté stránky.

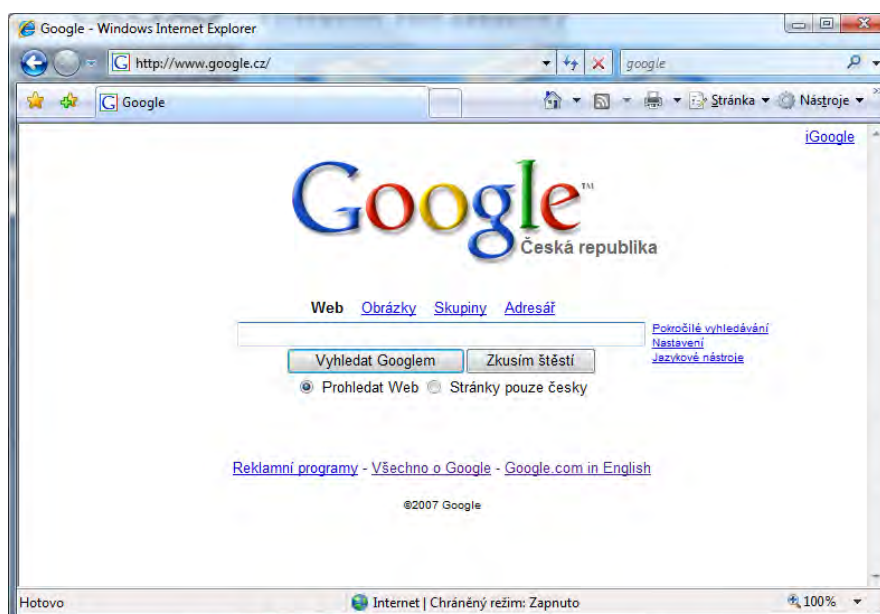
3.3.1 Vyhledávání ve fulltextovém vyhledávači

Pro zadávání dotazů fulltextovým vyhledávačem má každý z nich vytvořený specifický jazyk. Nemusí být vždy stejný, ale základní zásady platí pro všechny a liší se většinou pouze v jejich implementaci. Tento jazyk je v podstatě zjednodušený programovací jazyk, který stanovuje podmínku pro vyhledávání. Při stanovení vyhledávací fráze musíme vycházet i z daného jazyka (různé tvary, časy, synonyma apod.).

Správná vyhledávací fráze nemusí být výsledkem prvního pokusu. Někdy je požadovaná informace záludně schovaná a vmy musíme správné sousloví vhodně definovat.

3.3.1.1 Okno vyhledávače

Většina vyhledávačů působí velmi jednoduše. Nemají žádnou rozsáhlou grafiku ani další prvky, které rozptylují pozornost. Jejich předností je rychlost načítání a vyhledávání. Příkladem může být úvodní stránka vyhledávače Google.



Úvodní strana Google.cz

Na hlavní stránce je pouze vyhledávací formulář a několik odkazů na rozšířené vyhledávání. Nad polem pro zadání fráze je přepínač, kterým se určuje oblast hledání. Dvě tlačítka jsou určena pro odeslání dotazu. První z nich – Vyhledat Googlem zobrazí výsledky vyhledávání. Jistou alternativou je pak tlačítko Zkusím štěstí, které přímo zobrazí první nalezenou stránku.

3.3.1.1.1 Vyhledávací fráze

Volba správné vyhledávací fráze je klíčová pro nalezení požadované informace. Předpokládejme, že hledáme místa vhodná ke koupání na naší dovolené. První dotaz, který

vyzkoušíme, je „Koupání na dovolené“. Výsledkem záměrně chybně položeného dotazu je níže uvedená stránka.

Výsledky 1 - 10 z asi 1 400 000 na dotaz koupání na dovolené. (0,25 sekund)

Pod zopakovaným formulářem pro vyhledávání vidíme lištu s počtem výsledků klíčového slova. Jak je vidět vyhledaných stránek je asi 1 400 000, což je číslo velmi velké.

[koupání na dovolené | Moje-rodina.cz](#)

koupání na dovolené. Rubrika: Dovolená - Přihlašte se nebo zaregistrujte , abyste mohli psát komentáře | pošli tuto stránku | Zobrazit pro tisk ...
[www.moje-rodina.cz/dovolena/koupani-na-dovolene](#) - 36k - [Archiv](#) - [Podobné stránky](#)

Pod touto lištou jsou seřazeny jednotlivé vyhledané stránky. U každého záznamu je uveden titulek stránky a úryvek obsahu, který se týká vyhledané fráze.

Sponzorované odkazy

[Hotely a penziony](#)

nově, výběr zcela
podle vašeho přání na
[hotelselector.wlw.cz](#)

V levém sloupečku je pak zobrazen „sponzorovaný odkaz“, což není nic jiného než reklama, zobrazená na základě shody se zadaným klíčovým slovem inzerenta.

Goooooooooogle ►

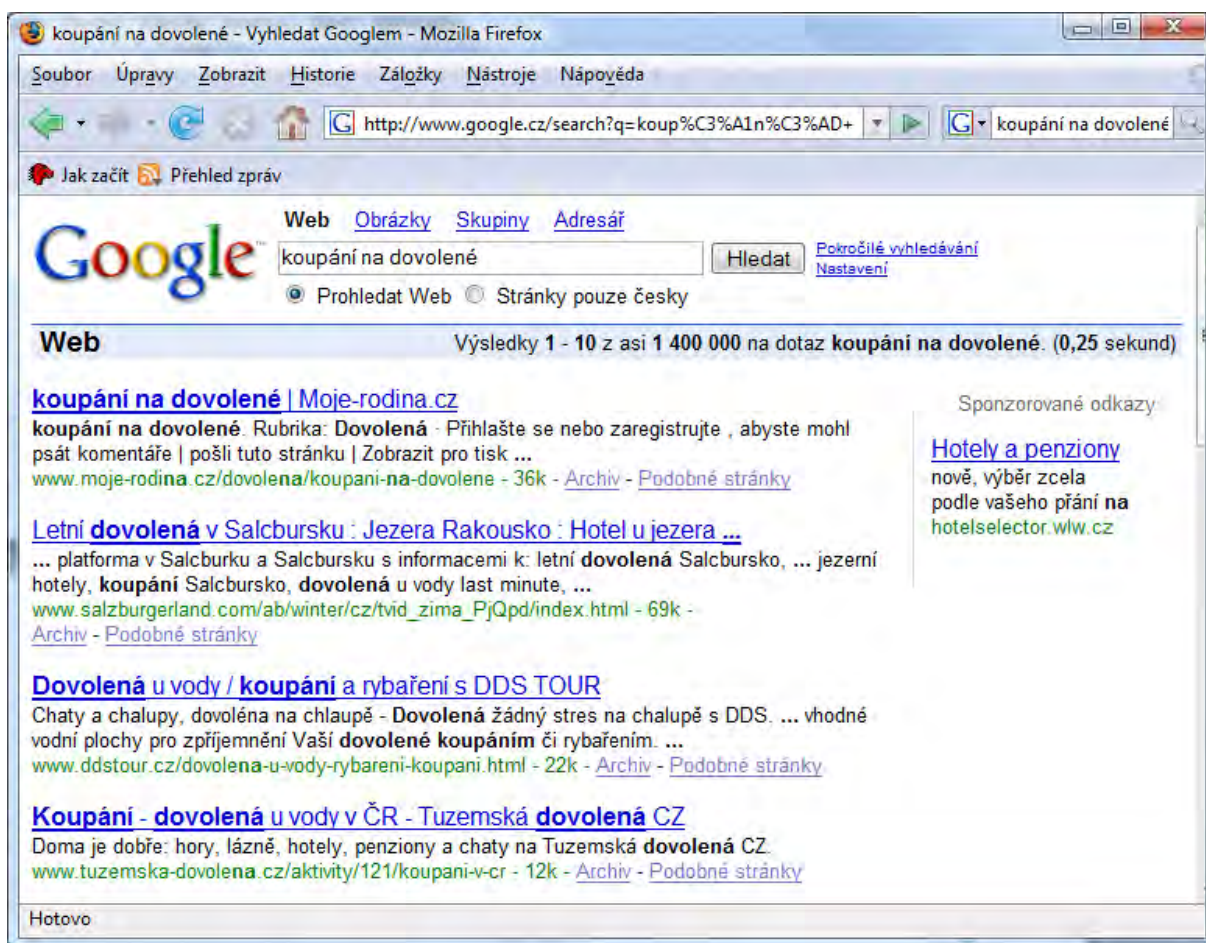
Stránka s výsledky: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [Další](#)

[Hledat ve výsledcích](#) | [Jazykové nástroje](#) | [Tipy pro vyhledávání](#)

[Google Home](#) - [Reklamní programy](#) - [Všechno o Google](#)

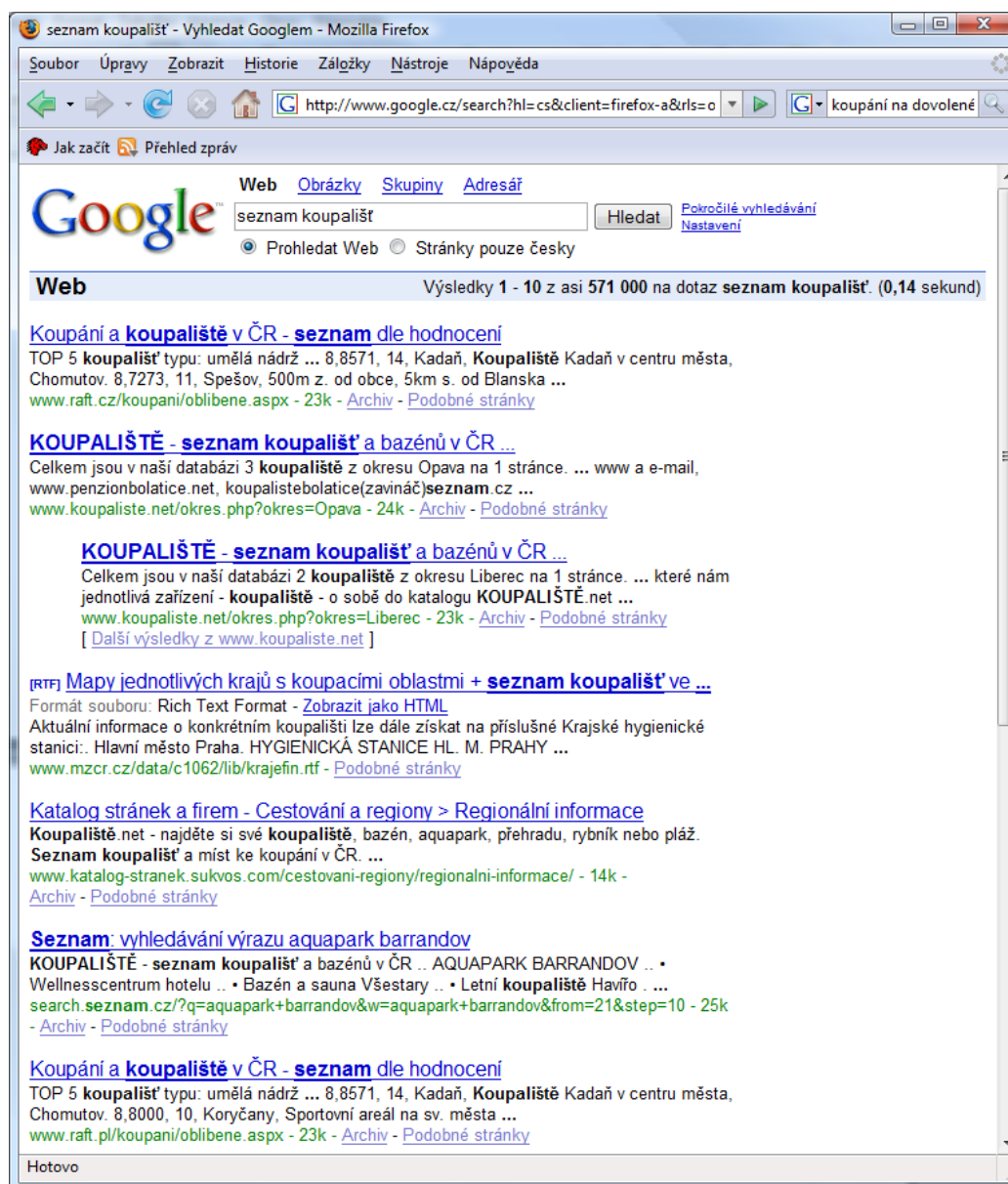
©2007 Google

Ve spodní části stránky je pak nástroj na listování nalezenými odkazy. Rovněž zde nalezneme možnost hledat ve výsledcích dotazu.



Výsledky vyhledání

Nicméně námi nalezené odkazy jsou v převážné většině pouze pojednání o koupání či dovolené. Je to pochopitelné, protože námi hledané informace jsou spíše strohé a méně textově obsáhlé než články z populárních časopisů. Upravíme tedy náš dotaz, respektive se zeptáme na to, co nás opravdu zajímá a to „seznam koupališť“.



Nyní je výsledek v podstatě dostačující, protože existuje velká pravděpodobnost, že námi nalezené stránky budou obsahovat potřebný nástroj pro vyhledávání, například s omezením pro náš kraj.

3.3.1.2 Pokročilé metody vyhledávání

V některých případech nám ale nalezený výsledek nemusí postačovat. Může se jednat o případ, kdy hledáme konkrétnější informace, například malou sjezdovku, která není v podobných vyhledávacích, nebo například otevírací dobu muzea, což může být informace neobsažená v nalezeném přehledu. Dalším případem může být vyhledávání informací o zboží. Zde je situace horší, protože existuje mnoho obchodů, které sice daný produkt nabízejí, ale jejich stránky obsahují pouze povrchní informaci z katalogu, nikoliv tedy potřebné doplňující informace.

V tomto případě můžeme použít rozšířených možností vyhledávačů. Jedná se především o fráze, vylučovací a podmiňovací značky. V případě, že v tomto „jazyku“ vyhledávače nejsme

zběhlí, můžeme použít formulář pro rozšířené vyhledávání, který nabízí uživatelsky přítulnou verzi této metody.

Formulář rozšířeného vyhledávání Google

Pokud chceme vyhledávání zrychlit, můžeme použít některé z následujících operátorů přímo ve formuláři pro vyhledávání.

AND (a také) - základní operátor, výsledek hledání zkrátí na stránky obsahující slova, mezi kterými se AND nachází.

OR (nebo) - vyhledá stránky, které obsahují alespoň jedno klíčové slovo (keyword), které OR spojuje.

NOT (ne) - vyhledá stránky, které neobsahují klíčové slovo za operátorem NOT.

Operátory AND a NOT lze nahradit znaky + a -.

NEAR/x (poblíž) - udává, v jaké vzdálenosti od sebe mohou být klíčová slova maximálně vzdálena. X určuje maximální vzdálenost (vzdálenost se „měří“ ve slovech a zapisuje se číslem).

FAR/x (daleko) - udává, v jaké vzdálenosti od sebe mohou být klíčová slova minimálně vzdálena. Dále je to jako u NEAR.

ADJ/x (z angl. adjacent - přilehlý) - podobné jako u NEAR a FAR, akorát nezáleží na pořadí slov a x označuje přesnou vzdálenost.

Meta words (meta-slova)

host:x nebo **site:x** - udává, na jaké adrese se bude vyhledávat. X je www adresa.

image:x - výsledek hledání bude zúžen na dokumenty obsahující zvolený obrázek. X je název obrázku.

link:x - vyhledá stránky obsahující odkaz na zvolenou adresu. X je www adresa.

anchor:x - vyhledá stránky obsahující v odkazech zvolené klíčové slovo. X je klíčové slovo.

domain:x - určuje doménu, na které se bude hledat.

url:x - určuje URL, na které se bude hledat. X je URL adresa.

title:x - vyhledá stránky s titulkem(hlavičkou) x.

applet:x - vyhledá stránky obsahující applety s názvem x.

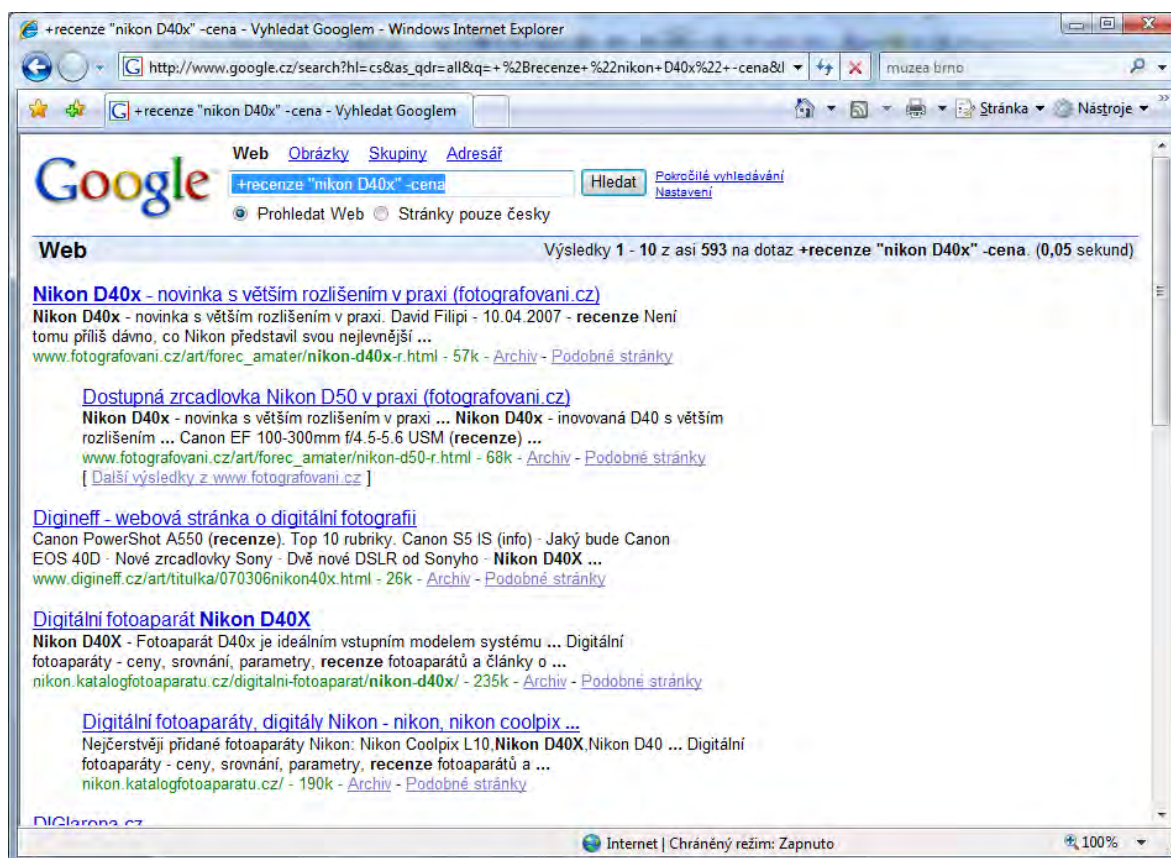
"x_y" - vyhledá všechny dokumenty obsahující frázi x_y (slova ve frázi oddělujte mezerou).

y?z - otazník nahrazuje jeden znak v hledaném výrazu.

uvw* - hvězdičkou se nahrazuje více znaků v hledaném výrazu .

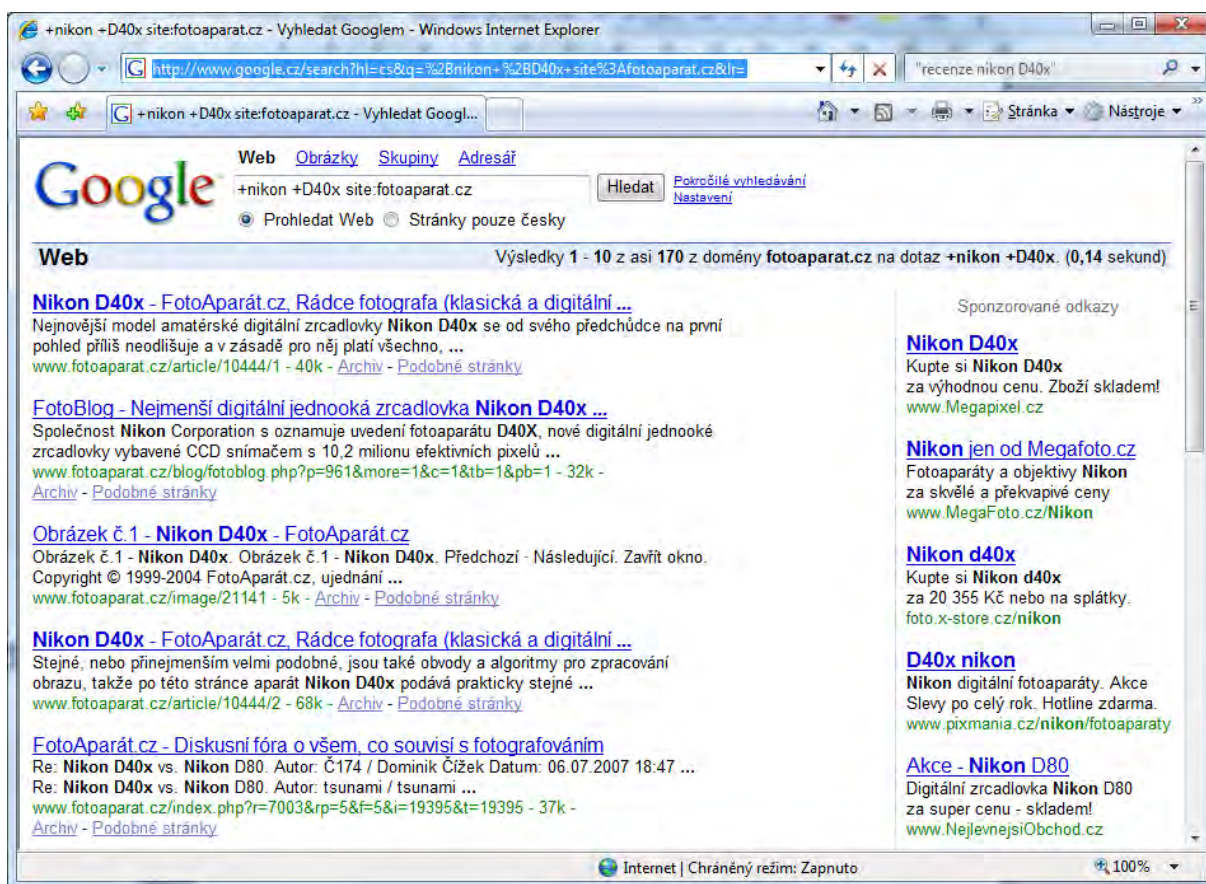
Ne všechny vyhledávače podporují všechny uvedené operátory, podrobné informace se dají nalézt v nápovědě každého vyhledávače.

Upřesňující vyhledávací fráze může vypadat takto: **+recenze "nikon D40x" -cena**. Pokud bychom zadali vyhledávací frázi **recenze nikon D40X** obdržíme 45 tisíc vyhledaných stránek. Omezení na frázi „Nikon D40X“ zajistíme vyřazení stránek nepojednávajících o tomto modelu, ale i o jiných modelech Nikonu. Podmíněným slovem + recenze zajistíme, aby v obsahu vyhledaných stránek bylo slovo recenze. Vyloučením slova **-cena** naopak eliminujeme obchody, které se na stránky snaží přilákat zákazníky hledající recenze produktů. Výsledkem je pak přijatelných 593 záznamů.



3.3.1.3 Prohledávání stránek

V některých případech poskytovatel internetového obsahu sice na svých stránkách má vyhledávací nástroj, ale ten je zpracovaný vysloveně žalostně. Může se stát, že dokonce vyhledá neexistující stránky nebo velmi nerelevantní informace. V tomto případě je možné použít vyhledávač i na prohledání těchto stránek. K vyhledávací frázi doplníme **site:x**, kde x je doména, kterou chceme použít. Například nás zajímají všechny zmínky o fotoaparátu Nikon D40x na serveru fotoaparát.cz, použitý dotaz bude vypadat **+nikon +D40x site:fotoaparát.cz**. Jak je vidět obdržíme 170 vyhledaných stránek. V případě neomezení dotazu na danou doménu bychom dostali pouze dva výsledky mezi odkazy na jiné stránky.



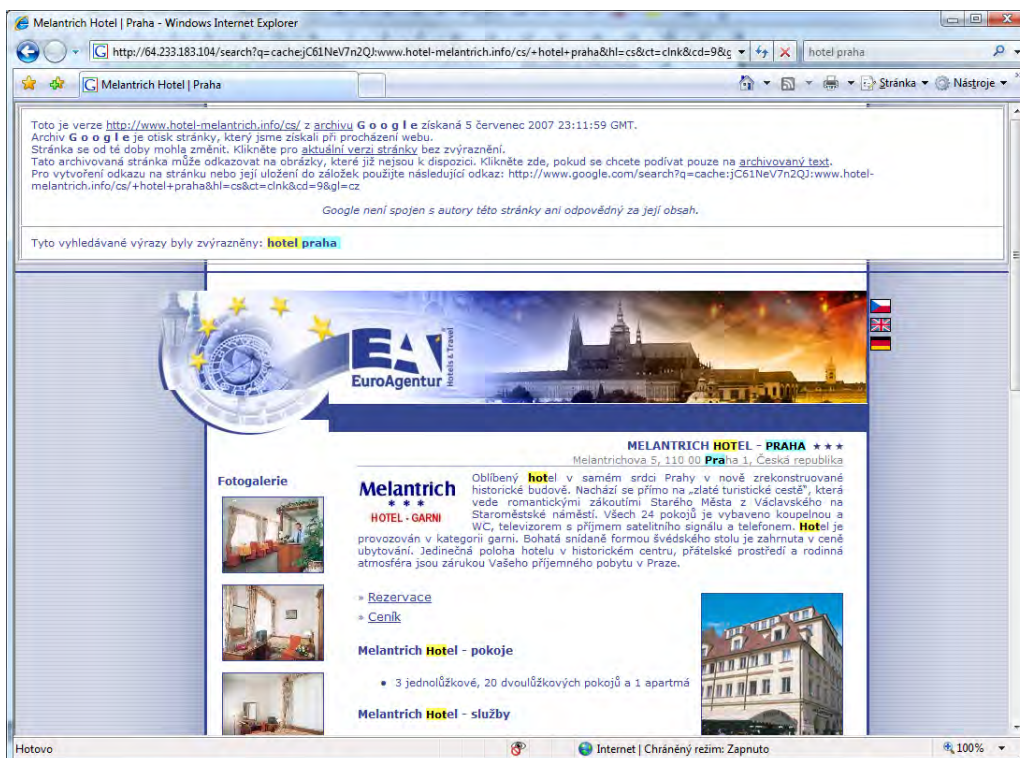
3.3.1.4 Další možnosti

Dalšími možnostmi vyhledávače Google je omezení vyhledaných stránek na jazyk, formát souboru, datum vytvoření webové stránky a umístění klíčového slova (v titulku stránky apod.).

3.3.1.5 Archiv

Některé webové prohlížeče rovněž zaindexované stránky archivují (např. Google). Archivu můžeme využít ve chvílích, kdy například webový server má výpadek a nebo již byla stránka stažena. Nicméně na přítomnost archivů bychom měli pamatovat i z druhé strany. Námi zveřejněný obsah je tímto v podstatě neodstranitelný.

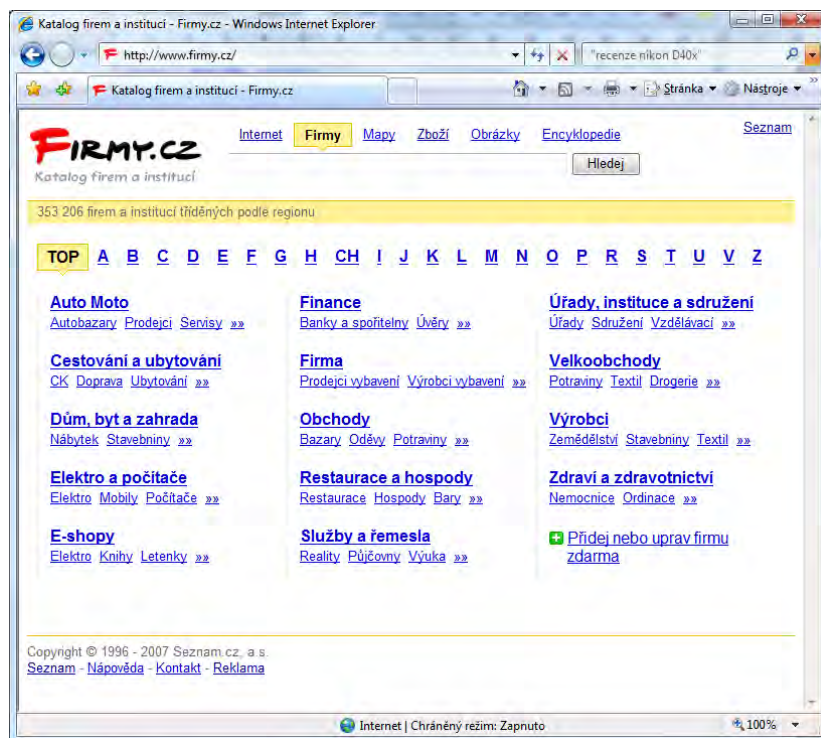
Archivní stránku zobrazíme kliknutím na odkaz archiv pod vyhledanými záznamy.



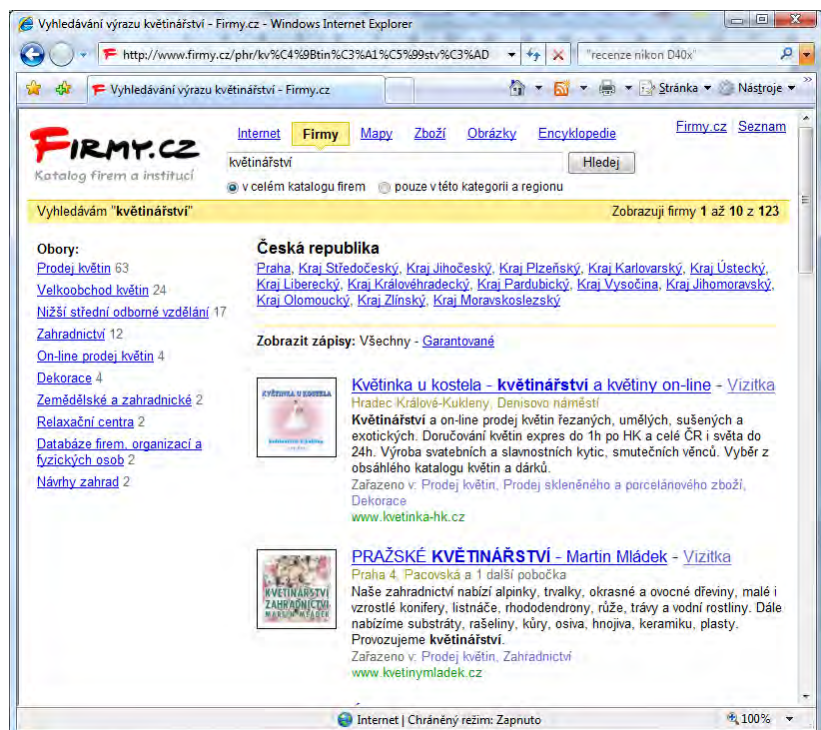
Stránka zobrazená z archivu

3.3.2 Vyhledávání v katalogu

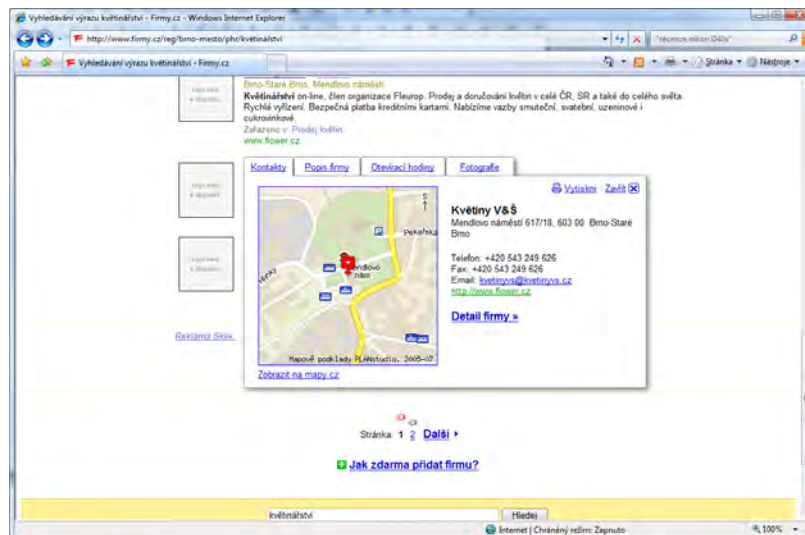
Vyhledávání v katalogu je poněkud specifičtější. Velkou váhu hraje zaměření katalogu. Vyhledávací mechanismus závisí na typu katalogu. Jako příklad uveďme katalog www.firmy.cz. Tento katalog patří Seznamu a obsahuje informace o firmách, jinde často nevyhledatelných. Úvodní stránka vyhledávače vypadá podobně jako u fulltextového, má ale navíc seznam sekcí. Jedná se v tomto případě o strukturu definovanou provozovatelem katalogu. Zaměstnanci provozovatele nebo sami uživatelé jej pak plní informacemi. Díky přesné struktuře je větší pravděpodobnost správného třídění dat, nicméně je problematičtější nalézt požadovanou kategorii. Proto katalog implementuje fulltextové vyhledávání v rámci katalogu.



Na dotaz „květinářství“ nám katalog našel přes 100 záznamů firem. Rovněž nám v levém sloupečku nabídne seznam relevantních kategorií. Díky ruční indexaci je rovněž možno zvolit kraj a poté i okres a město, ve kterém službu hledáme.



Výsledkem vyhledání může být jednak webová stránka, ale i vizitka s informacemi o firmě, kontakty, mapou a otevírací dobou.



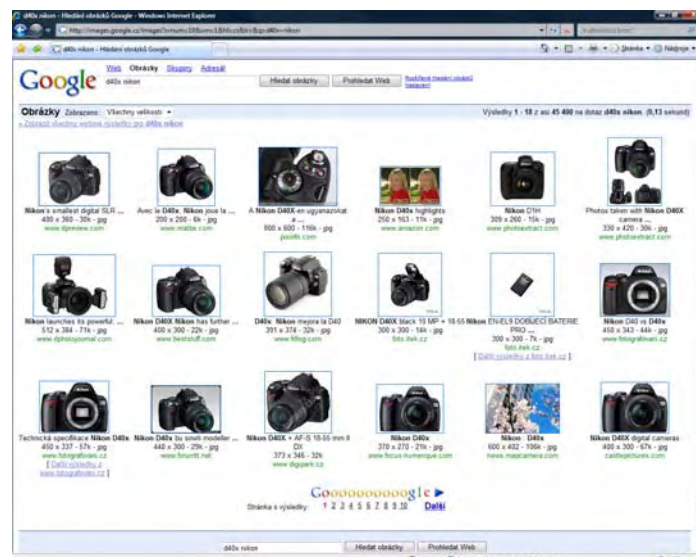
Jistým nedostatkem může být fakt, že v Brně nalezl katalog 11 květinářství, zatímco Google na dotaz „květinářství Brno“ 42 odkazů.

3.3.3 Vyhledávání obrázků

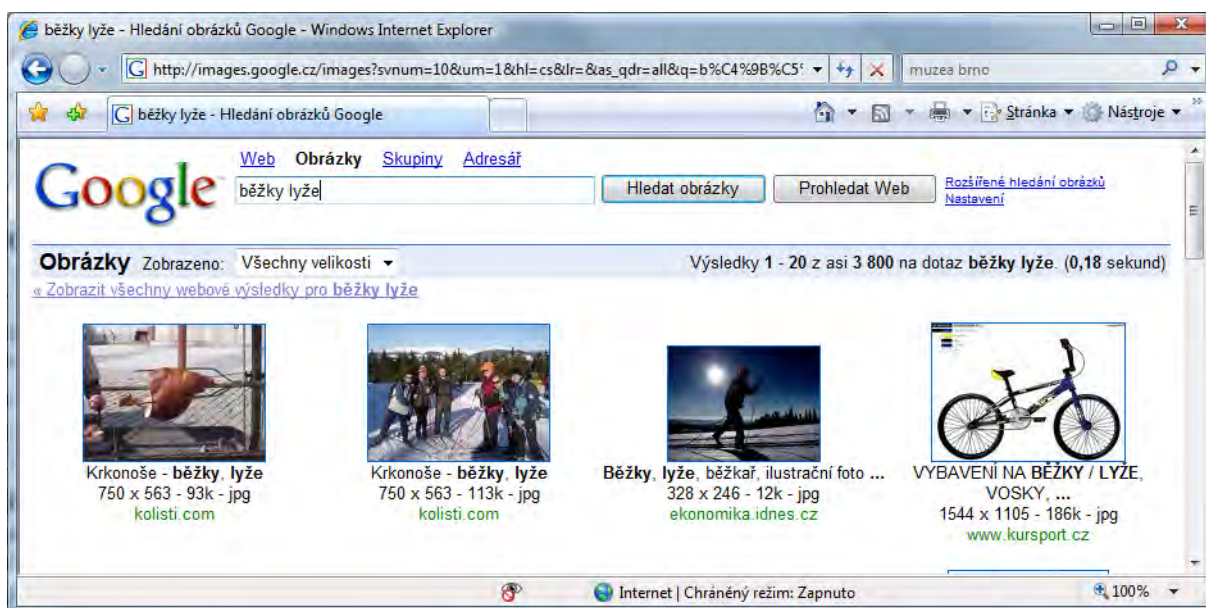
Specifickou oblastí je vyhledávání obrázků. Jak již bylo uvedeno, na rozdíl od textu, kde se vyhledávač může orientovat podle více informací, je u obrázků odkázán na okolní text nebo na některé jiné informace, například odkazy, alternativní text obrázku a jiné.

Pro vyhledávání obrázků slouží například images.google.cz. Základní vyhledávání funguje úplně stejně, problém nastává s relevancí odkazů.

Například na náš oblíbený dotaz D40x se mimo fotografií fotoaparátu objevují i fotografie pořízené tímto fotoaparátem.



Podstatně horších výsledků dosáhneme na obecnější slova nebo na méně frekventované a jednoznačné dotazy. Například na poměrně specifický dotaz „běžky lyže“ jsou mezi prvními čtyřmi výsledky dva špatné.



3.3.3.1 Rozšířené možnosti vyhledávání obrázků

Ve vyhledávání obrázků běžně můžeme zvýšit relevanci a použitelnost výsledků dotazu pouze o některé vlastnosti obrázku strojově rozpoznatelné. Jedná se o velikost, formát souboru a barevnou hloubku.

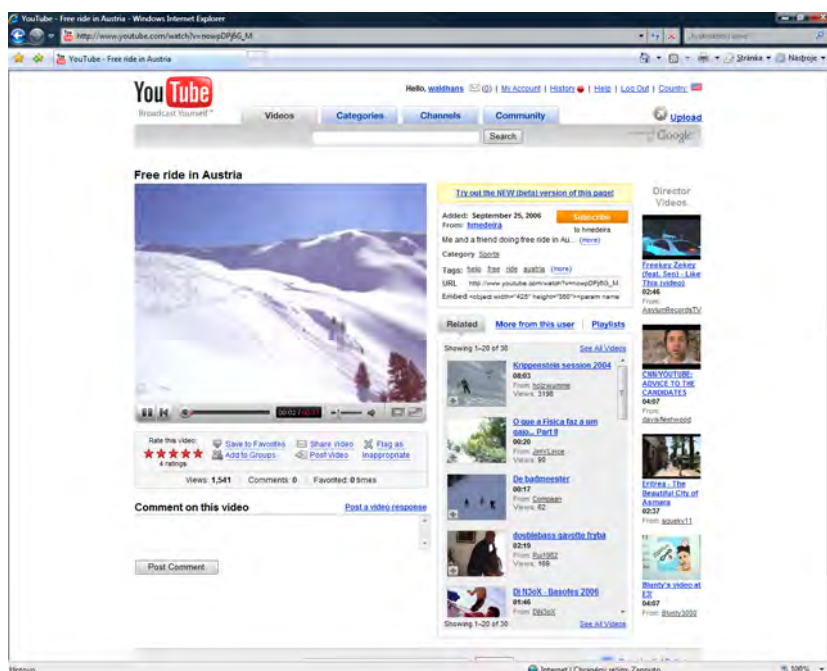
Novým parametrem u Google je možnost vyhledávání obličejů. Funkce zřejmě vychází z experimentování Google s automatickou analýzou obrazu.



Výsledky vyhledávání obrazu se zapnutým filtrem na obličej

3.3.4 Vyhledávání videa

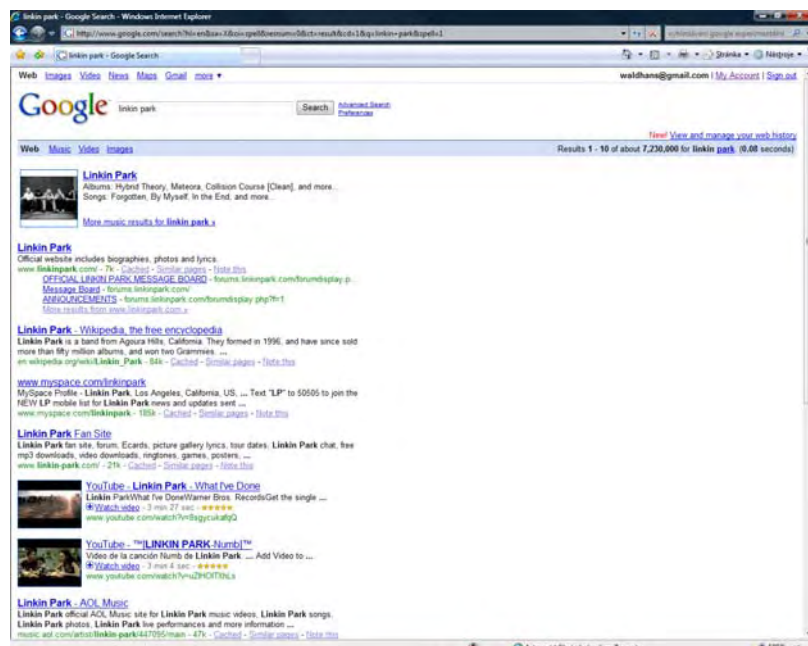
Zatímco u vyhledávání obrazu existuje alespoň nějaká šance na analýzu obrazu, u video obsahu není zatím ani náznak o něco takového. Proto se musíme spolehnout na správně fungující katalogy. Asi nejznámějším současným videoarchivem je Youtube. Stal se tak populárním a rozšířeným, že dokonce i Evropská komise si na něm zřídila vlastní kanál na kterém se snaží propagovat Evropskou unii. Vyhledávání nicméně probíhá pouze na základě informací, které autor o videu sdělí a rovněž podle diskuze pod videem.



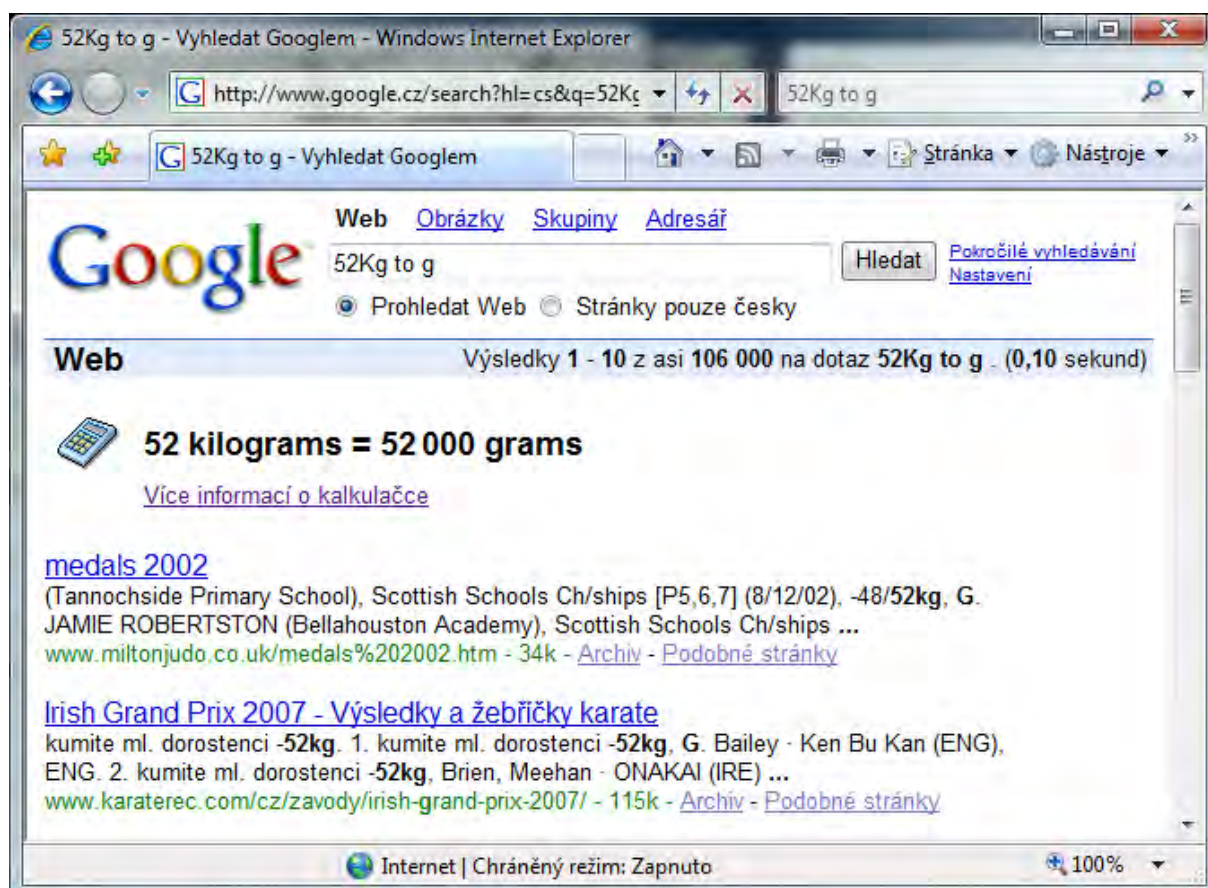
Formulář vyhledávání na Youtube.com

3.3.5 Trendy ve vyhledávání

Novým trendem ve vyhledávání, který zkouší Google je kombinovat vyhledávání různých zdrojů. Například po zadání hesla Linkin Park, což je populární americká skupina, se objeví ve výsledku vyhledávání kromě webových stránek i video a například diskografie.



Další zajímavou funkcí je kalkulačka, která dokonce umožňuje převod jednotek. Například dotazem **52Kg to g** dostaneme v prvním řádku správnou odpověď:

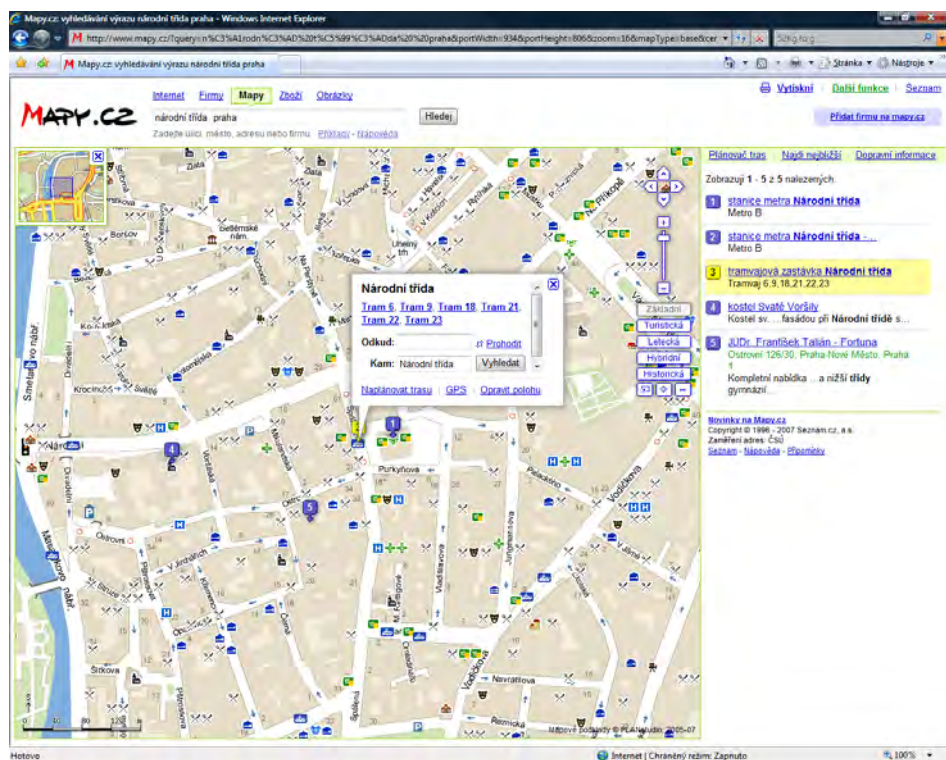


Experimentální funkce Google umožní například zobrazení časové osy pro hledané téma, vyhledávání oblíbenosti kapel a mnoho dalších.



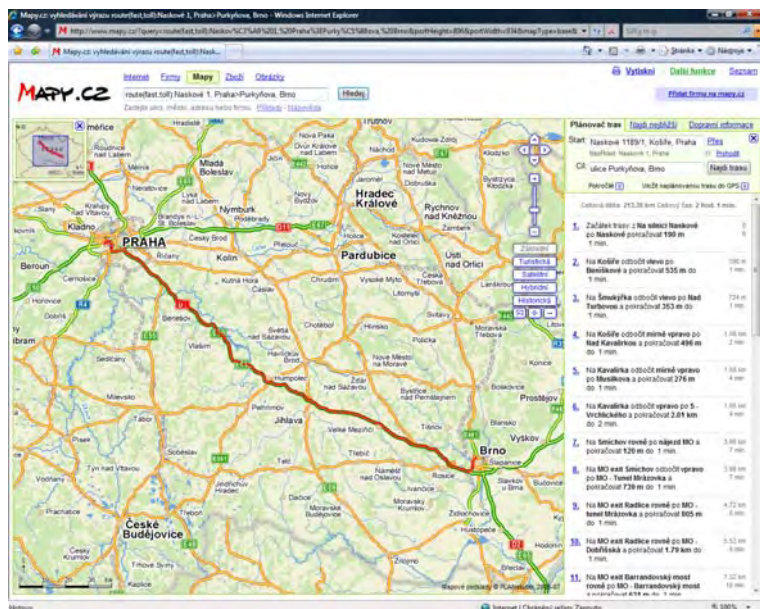
3.3.6 Mapy

Zajímavou službou, využitelnou obzvláště v cestovním ruchu, je vyhledávání v mapách. Pro Českou republiku je určena služba mapy.cz, nebo mapy.atlas.cz. Z celosvětového hlediska je nejpoužívanější služba Google maps. Výše uvedené služby umí vyhledávat i čísla domů. Některé mapy mají rovněž uživatelskou možnost doplnění obsahu k mapám a jejich použití na vlastních stránkách. V cestovním ruchu je využití ideální například k lokalizaci destinace, zprehlednění nabídky ubytování apod.



3.3.6.1 Plánovač tras

Plánovač tras je doplňková služba například vyhledávače mapy.cz. Na záložce stačí zadat počátek a konec trasy. Výsledná trasa je naplánována včetně délky trvání a případného exportu do některých GPS zařízení.



3.4 Optimalizace pro vyhledávače

SEO

(Search Engine Optimization,) je metodologie vytváření a upravování webových stránek takovým způsobem, aby byly ve výsledcích hledání v internetových vyhledávacích zobrazeny na nejlepších místech (tj. tam, kde je hledající hledají). Cílem je nalákat na vlastní stránky co nejvíce zákazníků (nezaměňujme s obecnějším „co nejvíce návštěvníků“).

3.4.1 Metody SEO

Cílem SEO je navrhnout stránku tak, aby byla pro relevantní dotazy hodnocena jako kvalitní a byla zobrazována na nejvhodnějších místech, protože většina uživatelů při hledání věnuje pozornost jen několika prvním odkazům. K tomuto cíli se používá mnoho různých technik, některé považované za správné a etické, jiné považované za nevhodné, neetické.

3.4.2 Etické metody

Etické metody SEO se snaží o vylepšení stránek samých pro všechny čtenáře, s tím, že se takovým vylepšením zvýší i kvalita stránek z pohledu prohledávacího robota. Mezi takové etické metody patří například:

3.4.2.1 Kvalitní a unikátní obsah

Stránka by měla mít kvalitní a unikátní obsah, pravidelně (v ideálním případě denně) aktualizovaný.

3.4.2.2 Správné používání doporučených značek (tagů)

Pokud bude tvůrce stránky používat na správných místech doporučené sémantické značky jazyka HTML či XHTML (například h1 pro nadpis první úrovně, em pro zdůraznění, strong pro silné zdůraznění), bude takovému textu vyhledávač přiřkládát větší význam. Jestliže webový tvůrce definuje nadpis pomocí velikosti písma (font size), nezjistí vyhledávací robot důležitost nadpisu.

3.4.2.3 Používání správného titulku (title)

Jestliže bude tvůrce uvádět v titulku konkrétní popis stránky (namísto obecných výrazů typu Úvodní stránka) či rovnou klíčová slova, získává stránka další významné plus.

3.4.2.4 Používání description, keywords

Vhodným používáním klíčových slov v meta description získá stránka v hodnocení robota další kladné body. Pokud se klíčová slova dále v obsahu stránky neopakují, nebo je jich moc, ztrácí na prioritě. Meta description je použit u popisu stránky ve výsledcích vyhledávání (SERP), použití vhodného textu může tedy zaujmout potenciálního návštěvníka a zvýšit míru prokliků.

3.4.2.5 Budování zpětných odkazů

Pokud je na stránku odkazováno z jiných webů, zvyšuje se její hodnocení (v případě Google je to PageRank, v případě Jyxo.cz je to jyxorank). Toho se dříve zneužívalo pro tzv. google bomby, kdy se pomocí velkého množství odkazů s určitými klíčovými slovy dostala na první

místo ve výsledcích hledání nesouvisející stránka. Příkladem Google bomby byla klíčová slova ostudné selhání (v angličtině), která na prvních místech zobrazila odkaz na web amerického prezidenta George Bushe. Podle posledních pozorování si nyní Google kontroluje, jestli je na stránkách související téma. Jestliže bude například ze stránky o chovu velryb odkazováno na web o pěstování muškátů, nebude mít odkaz takovou váhu. Kvalitní obsah zvyšuje šanci, že bude na web odkazováno z jiných zdrojů.

3.4.2.6 Krátká a neměnná URL adresa

Uvedení krátké URL adresy podpoří zájem ostatních uživatelů o odkazování na stránku. Příliš dlouhá a lidem nesrozumitelná URL adresa naopak od odkazování spíše odrazuje. Přítomnost klíčového slova v URL adrese může u některých vyhledávacích strojů zvýšit umístění stránky ve výsledcích hledání těchto slov. Naopak se nedoporučuje používat v URL parametr id. Požadavek na neměnnost adres vyplývá z fungování vyhledávacích strojů i z požadavku na budování zpětných odkazů.

Špatná adresa z hlediska SEO:

<http://cestovka.cz/kat.php?id=432&what=B603AA60CC16206AC1257054003EA0EB&kat=jjdd>

Dobrá adresa z hlediska SEO:

<http://cestovka.cz/zajezdy/plachetnice-chorvatsko-2007>

obsahuje položky reprezentující strukturu katalogu

3.4.3 Neetické metody SEO, spam

Kromě etických metod se některé webové stránky snaží vylepšit svou pozici ve vyhledávacích podvodnými metodami. Proti takovým technikám se vyhledávače obvykle snaží bránit stránkám, které takové techniky používají, hrozí (při odhalení) postih a výrazný propad ve výsledcích, avšak přinejmenším krátkodobě takové techniky mohou být úspěšné.

Mezi typické neetické metody patří například:

3.4.3.1 Spam

Jelikož stránka, na kterou vede velké množství zpětných odkazů, je obecně považována za lepší, mohou se tvůrci stránek snažit zvýšit počet takových odkazů tím, že budou odkazy na své stránky umisťovat do internetových diskusí, na wiki, do komentářů k příspěvkům v blozích apod., aniž by tyto odkazy byly na těchto místech vítány či tam měly svoje místo.

3.4.3.2 Oklamání robotů

Internetové vyhledávače používají pro procházení internetu specializované programy, tzv. roboty. Tvůrce stránky se může pokusit o to, zobrazovat robotům jiný obsah než ostatním návštěvníkům a tím se dostat na přední pozice i se stránkami, se kterými by bez podvodu neuspěl.

3.4.3.3 Skrytý text

Vyhledávače berou v potaz při řazení výsledku, mimo jiné, frekvenci klíčových slov, frází. Tvůrce stránky si tuto frekvenci může uměle zvýšit tím, že dá na stránku text, který obsahuje

několikrát různá klíčová slova, a ten následně skryje, nebo ho nechá zobrazit velice malým písmem, aby normálního návštěvníka stránky nerušil.

3.4.4 Google bomba

Jako Google bomba se označuje jistý způsob „zneužití“ vlastností internetového vyhledávače Google (případně, v přeneseném významu i jiného), po kterém je výsledkem vyhledání nějaké fráze webová stránka, která danou frázi neobsahuje, ale útočník na ni tak odkazuje, obvykle s humorným nebo politicko-kritickým úmyslem.

3.4.4.1 Princip útoku

Vyhledávací algoritmus Googlu přihlíží nejen k obsahu stránek, ale i k tomu, jakým textem na ně odkazují jiné stránky, tj. de facto jak cílovou stránku popisují jiní lidé. Pokud např. nějaká stránka obsahuje odkaz na <http://www.wikipedia.org> s textem „velká encyklopedie“, při vyhledávání termínu „velká encyklopedie“ se bude uvažovat i stránka <http://www.wikipedia.org>, i pokud se na ní tento termín nenachází. Tento postup má zvýšit relevanci výsledků, ale lze ho zneužít.

Pokud mnoho stránek obsahuje odkaz na cílovou stránku s nějakým urážlivým textem (např. hlupák), bude Google mezi výsledky vyhledávání tohoto textu uvádět i cílovou stránku. Je-li odkazů mnoho, může se cílová stránka ocitnout na prvním místě ve výsledcích, což je obvykle cílem Google bomby. Takový výsledek lze zjednodušeně prezentovat jako „Google si myslí, že tato stránka pojednává o hlupákovi“ (správnější je ovšem „mnoho stránek uvádí, že tato stránka pojednává o hlupákovi“).

Princip objevil Adam Mathes v dubnu 2001, když mu Google na dotaz „internet rockstar“ vrátil stránku amerického informatika Bena Browna, přestože toto spojení neobsahovala. Mathes svůj nápad brzy vyzkoušel a za pomoci dalších bloggerů docílil, že stránka jeho kamaráda Andy Pressmana byla na prvním místě po dotazu „talentless hack“ (netalentovaný pisálek).

Nicméně první Google bomba zmiňovaná v tisku je „more evil than Satan“ (zlejší než Satan), která vedla na stránku firmy Microsoft už v roce 1999. Pravděpodobně šlo ale jen o náhodu, které si všimli uživatelé Googlu, nikoliv cílený útok.

3.4.4.2 Opatření

Na oficiálním blogu Google bylo oznámeno, že vývojáři přijali proti Google bombám opatření. Pokusili se snížit jejich efekt implementací obecného algoritmu, který by měl fungovat v různých jazycích. Webmaster zdůvodnil tento krok tím, že bomby byly často mylně interpretovány jako „názor Googlu“.

Google bomby obvykle nejsou příliš trvanlivé. Pokud jsou úspěšné a vejdou v širší známost, píše se o nich na řadě zpravodajských webů s vysokým PageRankem, jež původní cíl bomby vytlačí z předních míst. Mimo to pořadí výsledků všech vyhledávání běžně osciluje podle toho, jak Google přepočítává PageRank při aktualizacích svého indexu a pozměňuje parametry algoritmu; speciálně v únoru 2005 došlo ke změně, která většinu starších Google bomb oslabila.

3.5 Právní aspekty vyhledávačů

Internet jako médium, vzhledem ke své podstatě, umožňuje rovněž šíření nelegálního obsahu, ať už se jedná o duševní vlastnictví, nebo přímo kriminální obsah. Hlavně v poslední době vyvstává problém pro vyhledávače s jejich statutem. Vyhledávač jako takový nelegální obsah neposkytuje, ale pozve na něj „ukazuje“. Problém nastává v případě že se jedná o nelegální obsah, obsah chráněný autorským zákonem apod. Posledním případem je zatím neúspěšná žaloba na Google za digitalizaci knih. Jedná se o vyhledávání v knihách, kdy v případě vypršení autorské ochrany Google zobrazí knihu celou, v opačném případě jen úryvek. To se nelíbí knižním vydavatelstvím. Rovněž pravidelně řeší vyhledávače spory o audiovizuální obsah. Některé vyhrávají vyhledávače, ale už se objevují první vlaštovky, kdy nebyl provozovatel odsouzen za trestnou činnost, ale za její umožnění. Některé případy působí velmi kuriózně, například žaloba belgických novin na Google za zaindexování a zveřejňování jejich stránek.

Žaloba Preispiraten vs. Google

Žalobce provozuje internetové stránky na adrese www.preispiraten.de, prostřednictvím kterých nabízí ke stažení software určený k automatizovanému vyhledávání nejvýhodnějších podmínek nákupu zboží či služeb na různých internetových obchodech či aukcích. Podstatné je, že má zapsanu ochrannou známku Preispiraten (v překladu cenoví piráti). Konkurent žalobce začal na stránkách preisserver.de provozovat prakticky obdobnou službu, navíc použil zaměnitelný webdesign (který záhy na základě výzev žalobce změnil). Žalobce nakonec dosáhl kvůli porušení autorských práv a práv k ochranné známce ukončení této činnosti a převedení domény na žalobce.

Žalobce si dále všimnul, že po zadání slova „preispiraten“ do internetového vyhledávače Google se prostřednictvím systému AdWords zobrazí reklama na server konkurenta (preisserver.de). Zažaloval tedy přímo společnost Google a požadoval náhradu škody a zdržení se jakéhokoliv užívání ochranné známky „preispiraten“ jako klíčového slova ve zmíněném systému AdWords.

Soud však žalobu zamítnul, když konstatoval, že provozovatel systému AdWords nenese odpovědnost za obsah reklamního systému, který je určen k tomu, aby si klíčová slova zadávali přímo jeho uživatelé – inzerenti. Jeho odpovědnost by připadala v úvahu pouze tehdy, pokud by věděl o konkrétním porušení práv k ochranné známce, což se nestalo. V žádném případě nelze provozovatelům obdobných služeb ukládat povinnost proaktivního vyhledávání a ověřování uloženého obsahu.

Soud dále konstatoval, že žalobce podniká pouze v omezeném segmentu služeb a rovněž jeho ochranná známka je zapsaná pro omezený okruh zboží a služeb (obdobně se v ČR zapisují tzv. třídy ochranných známek) – tato skutečnost mu neumožňuje požadovat po žalovaném zdržení se jakéhokoliv užití pojmu totožného s ochrannou známkou žalobce.

Dle soudu nešlo ani o porušení práv k ochranné známce, protože v textu reklamy samotné se chráněný pojem „preispiraten“ nezobrazil a ani z textu nevyplývalo žádné spojení či souvislost s žalobcem. Navíc dle soudu nejde o obdobu metatagu, protože je reklamní text zcela jasně označen jako reklama a nehrozí tak „zmatení“ uživatele a jeho omyl, jako v případě zneužití metatagu.

Případy „Bourse des vols“, a „Impuls“

Oba v nadpise uvedené případy mají obdobnou skutkovou podstatu jako shora zmíněný případ „preispiraten“, nicméně soudy zde došly k opačným závěrům. Bourse des vols lze

přeložit jako „burza letů“ či „letová burza“ a francouzská cestovní kancelář, která má tento pojem zapsán jako ochrannou známku, vyhrála ve Francii soud I. instance i soud odvolací a vysoudila náhradu škody 75 000 EUR. Důvodem bylo právě umístění zmíněného pojmu do reklamního systému AdWords leteckou společností EasyJet.

Případ „Impuls“ se, co se týče metatagů, dostal až před německý Spolkový soud (obdoba českého Nejvyššího soudu). Společnost Impuls Mediamarketing GmbH, která nabízí srovnání zdravotních pojišťoven a poradenské služby v této oblasti, se cítila poškozena svým konkurentem, který si opět zadal do systému AdWords zapsanou ochrannou známku „Impuls“. Zažalovala však nikoliv provozovatele systému, nýbrž svého konkurenta o to, aby se zdržel využívání uvedeného klíčového slova v systému AdWords.

Soud (I. instance i odvolací soud) v tomto případě dal zapravdu žalobci a výslovně uvedl, že klíčová slova v AdWords je třeba posuzovat stejně, jako metatagy. Soud zdůraznil, že z jednání žalovaného vzniká nebezpečí záměny. K tomu, aby došlo k porušení práv k ochranné známce přitom není nutné, aby byla v reklamním textu přímo uvedena. Postačí totiž, že zadáním klíčového slova došlo k tomu, že uživatel je odkazován ke konkurenčnímu subjektu, než je ten, který si ve vyhledávači jmenovitě zadal. Soud tedy záležitost porušování práv k ochranným známkám posoudil zcela opačně než soud v případě „preispiraten“

Případ „Louis Vuitton“

Společnosti Louis Vuitton (výrobce luxusních módních doplňků) zažalovala opět přímo Google, a sice kvůli falšování zboží a nekalé soutěži.

Specifikum tohoto případu spočívá v tom, že se zde po zadání ochranné známky „Louis Vuitton“ popř. jiných jím zapsaných a známých známek objevila mj. reklama na prodejce imitací výrobků této značky. Ze strany prodejců a výrobců imitací jde o zcela jasné porušení práv k ochranné známce a případně též dalšího duševního vlastnictví. Žalovaným zde byl opět přímo Google a žalobce poukazoval na skutečnost, že reklamní systém AdWords při zadávání reklamy nabízel v rámci tzv. optimalizace seznam klíčových slov, přičemž po zadání klíčového slova „Vuitton“ systém nabídl inzerentovi mj. slovní spojení „Louis Vuitton repliky“, „Louis Vuitton kopie“ nebo „Fake Louis Vuitton bags“.

Odvolací soud v létě minulého roku potvrdil a ještě zpřísnil rozhodnutí soudu první instance a odsoudil Google k náhradě škody ve výši 300 000 EUR a náhradě nákladů řízení ve výši 60 000 EUR. V odůvodnění uvedl, že vědomě užíval ochranné známky k nabídce produktů, které nebyly vyrobeny jejím majitelem. Tím dosahoval zisku k přímé škodě společnosti Louis Vuitton.

Soud Google dále uložil, aby vyvinula systém, který takovým případům bude předcházet. Přitom soud zmínil chování společnosti Google v Číně, kde na základě pokynů tamní vlády aplikoval systém, který vylučuje z vyhledávače politicky citlivé pojmy.

Výše uvedené případy ukazují, že některé spory jsou velice neobvyklé a časem jich bude přibývat. Někdy se ale nelze ubránit pocitu, že se někdo chce přizivit na úspěchu druhých.

Slovníček

Zpětný odkaz

Jako zpětný odkaz (anglicky backlink, inbound link) se označuje každý odkaz mířící na danou stránku z jiné stránky. Primární funkcí zpětných odkazů především přivádět na web nové návštěvníky. Zpětné odkazy zvyšují viditelnost webu.

URL

Celým názvem Uniform Resource Locator („jednotný lokátor zdrojů“) je řetězec znaků s definovanou strukturou, který slouží k přesné specifikaci umístění zdrojů informací (ve smyslu dokument nebo služba) na internetu.

Hypertext

je informační systém, který zobrazuje informace v textu, který obsahuje návěští odkazující na upřesnění nebo zdroje uváděných informací tzv. hyperlinky neboli česky (hypertextové) odkazy. Rovněž odkazuje i na jiné informace v systému a umožňuje snadné publikování, údržbu a vyhledávání těchto informací. Nejznámějším takovým systémem je World Wide Web.

Konverzní poměr

je často označován jako míra konverze, označuje statistickou pravděpodobnost, že se z návštěvníka stránky stane zákazník.

Jednotky množství informace v informatice

Bit

1 bit: v „digitální abecedě“ se jedno písmeno skládá z osmi bitů

Bajt (8 bitů)

Kilobajt (1000 bajtů nebo 10 na 3 bajtů)

Megabajt (1000 kilobajtů nebo 10 na 6 bajtů)

Gigabajt (1000 megabajtů nebo 10 na 9 na bajtů)

Terabajt (1000 gigabajtů nebo 10 na 12 bajtů)

Petabajt (1000 terabajtů nebo 10 na 15 bajtů)

Exabajt (1000 petabajtů nebo 10 na 18 bajtů)

Zetabajt 1 000 000 000 000 000 000 000 nebo 10 na 21 bajtů

Jotabajt 1 000 000 000 000 000 000 000 000 nebo 10 na 24 bajtů

Použitá literatura

Hlavenka, Jiří: Mistrovství ve vyhledávání na Internetu, Nakladatelství Computer Press, a.s., 2002, ISBN: 80-7226-759-0

Hlavenka, Jiří: Jak na počítač - Vyhledávání na Internetu, Nakladatelství Computer Press, a.s., 2005

Šebesta, Ondřej: Hledáme na internetu, Grada 2001

Iskra, Jiří: Google - Vyhledávání, Gmail, Google Talk a další služby, Nakladatelství Computer Press, a.s. 2006, ISBN: 80-251-1043-5

Lapáček, Jiří; Klíma, Miroslav: Jak najít na internetu, Nakladatelství Computer Press, a.s., 2002

Procházka, David: Hledáme na internetu, Grada Publishing 2007, ISBN: 80-247-1471-4

www.pcsvet.cz

www.imation.cz

www.wikipedia.org

www.cesnet.cz

www.pravoit.cz

www.jakpsatweb

4 MODUL 4 - ZPŮSOBY VYHLEDÁVÁNÍ V INTERNÍCH A KOMBINOVANÝCH DATOVÝCH ÚLOŽIŠTÍCH



Informace nezbytné pro každodenní práci většiny organizací jsou rozprostřeny přes velké množství registrů, systémů a institucí. Tyto informace jsou uloženy v heterogenních úložištích a velmi často v nestrukturované podobě. Jejich sjednocení do společného datového skladu a struktury je nereálné. Řešením může být vybudování systému založeného na inteligentních vyhledávacích nástrojích, které umožňují propojení různorodých zdrojů dat do jednotného informačního prostředí. Toto prostředí dovoluje všem pracovníkům vyhledávat a spojovat data ve vlastní organizaci a jejích datových úložištích, v prostředí virtuálních privátních sítí spolupracujících institucí i ve veřejných datových zdrojích privátních subjektů. Po posouzení relevance a bezpečnosti informací lze vybrané údaje a statistiky publikovat i široké veřejnosti. Systémy tohoto druhu umožňují též kladení dotazů a třídění informací ve více jazycích, což má velký význam zejména v okamžiku, kdy informace pocházejí z nadnárodních informačních kanálů nebo jsou kombinací zdrojů z více jazykových, kulturních či legislativních prostředí.

V této části bychom Vás chtěli seznámit s filozofií systémů pro vytěžování dat a poskytnout Vám návod, jak je již dnes, **bez nutnosti investic**, začít používat. V první fázi jde o data uložená ve Vašem PC, Vaší poště a na Internetu. Poskytneme Vám ale také vodítko, jak rozšířit možnosti efektivního vyhledávání na celou Vaši organizaci či systém institucí (taková infrastruktura již vyžaduje placené verze software a hardware).

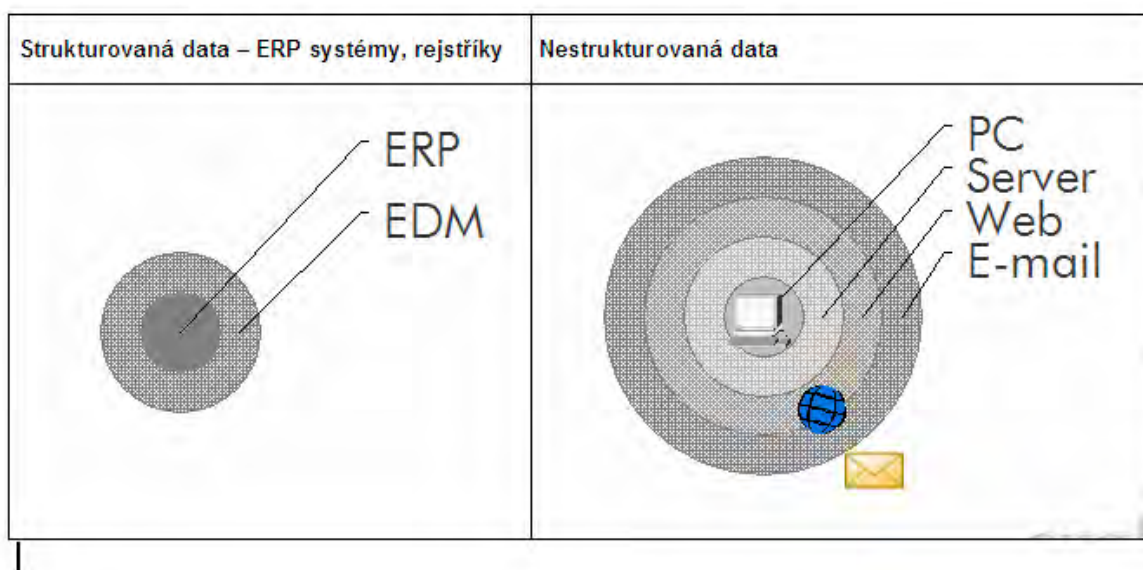
4.1 Uložení informací

Dle četných výzkumů stráví většina z nás 35% své pracovní doby vyhledáváním informací a 40% z nás pravidelně nemůže najít určité konkrétní údaje, přestože víme, že už jsme se s nimi při své práci někde setkali.

POZNÁMKA:

- Požadované informace jsou ponejvíce nestrukturované
- Denně je odesláno více než 210 milionů e-mailů (IDC Research)
- 45% relevantních informací je uloženo v přílohách (IDC Research)
- Ročně přibývá 5 exabytů (30 000 národních knihoven) dat na pevných discích

Informace, které hledáme nejčastěji bývají uloženy na některém z úložišť dle následujícího schématu:



Současný stav ve většině organizací je takový, že pro vyhledání informací v každé z domén dle výše uvedeného schématu se používá specializovaný nástroj či metoda. V systému ERP to je vnitřní funkcionality prohlížení údajů, v poště (nejčastěji Outlook) nabídka Nástroje – Najít, na vlastních discích prostředky operačního systému (Windows Explorer) a na Internetu některý z vyhledávačů (nejčastěji Google). Velmi často uživatel hledá informaci ve všech těchto zdrojích najednou a výsledek manuálně skládá do požadované podoby výsledku.

Velkým problémem je bezpečnost a důvěryhodnost dat. I pouhá existence mnohých údajů zůstává uživateli skryta za hranicemi specializovaných systémů, přestože nepodléhají žádným systémům utajení, a naopak informace utajované jsou dostupné na sdílených úložištích serverů.

Tento stav je vzhledem k růstu množství dat v organizacích i Internetu neudržitelný a má velmi negativní vliv na produktivitu i kvalitu práce.

V následující tabulce je uveden přehled **některých** agend typické organizace a běžný způsob uložení dat, která se v daném oddělení zpracovávají.

Oddělení	Druh agendy	Typické úložiště
----------	-------------	------------------

Vedení		Správní agenda, spisová agenda, metodické pokyny....	Souborový server, dokumenty Word, Excel, Acrobat pdf, snímky dokumentů v souborovém či databázovém úložišti.
Hospodářská správa		Správa budov, majetku	Souborový server, Word Excel, databázové systémy: Oracle, MSSQL, dbf (Foxpro)
	Podatelna	Pošta včetně elektronické	Poštovní server, Outlook, Fedora, Scany dokumentů, Word, Excel
	Správa budov	Údržba, pronájmy, zakázky	Databázové soubory, Word, Excel, pdf, specializovaný software
	Autoprovoz	Provoz parku, leasing	Specializovaný software
	Technická správa	Ostraha, stravování, úklid.....	Tato oblast je velmi členitá a je obsluhována velkou řadou standardního i proprietárního software
Finanční správa	Rozpočet Účtárny	Účetní agenda	ERP systémy
Evidence	Zákonem stanovené evidence a registry	Nejrůznější registry, které jsou vedeny na základě zákonů a nařízení	Specializované moduly, databázové soubory, Excel.

Z uvedeného je tedy zřejmé, že k účinnému hledání požadovaných informací již nestačí internetové vyhledávače a více či méně manuální práce s prostředky operačního systému. Pro přeměnu nesourodých skladišť dat v užitečné informace je třeba využít novou techniku vyhledávání a organizace dat.

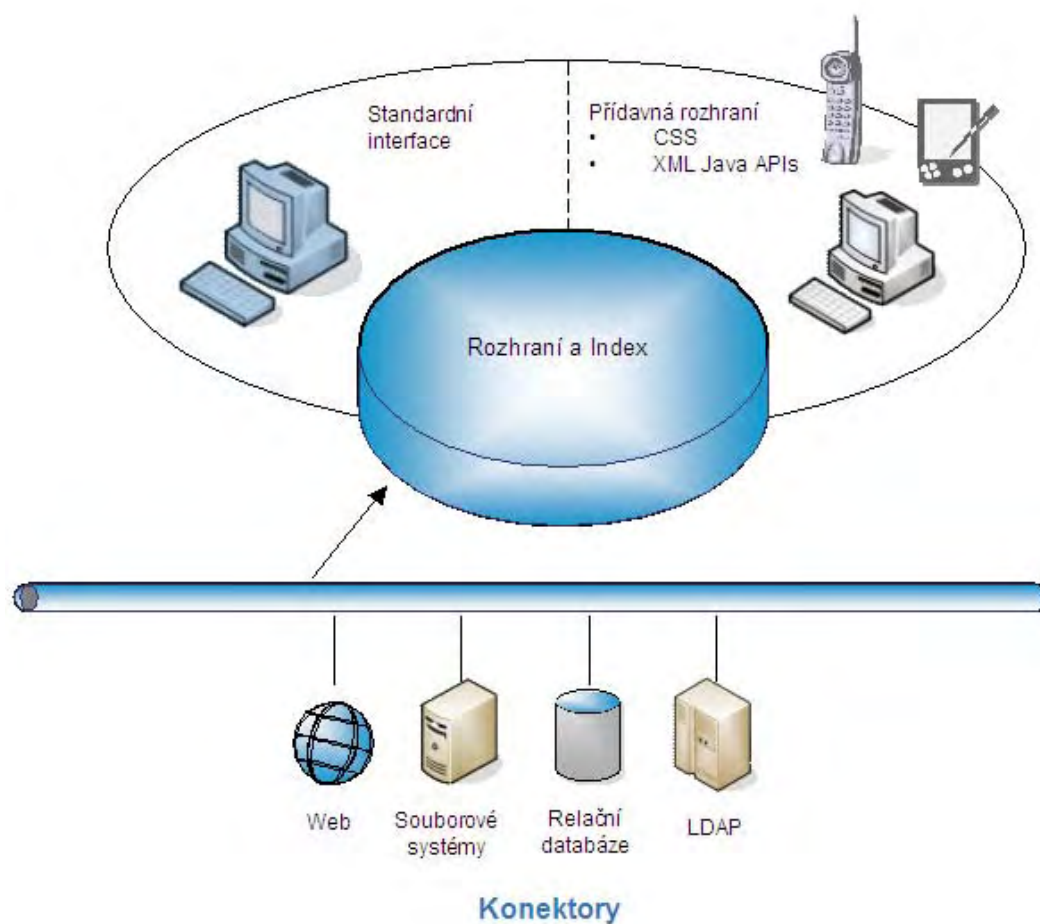
4.2 Vlastnosti aplikací pro komplexní vytěžování informací

Na aplikaci, která má být opravdovým prostředkem pro vytěžování informací jsou kladeny mnohem větší nároky než na prostý vyhledávač. Takový systém:

- Jednoduše a efektivně zprostředkovává všechny druhy vyhledávání – od dat uložených v osobním počítači, přes data organizace až po Internet a další veřejné zdroje.
- Uspořádává výsledky strukturovaně dle preferencí uživatelů.
- Respektuje zásady zabezpečení nalezených informací.
- Podporuje hledání technikou podobných výrazů a dovoluje definovat vztahy mezi hledanými údaji (například pomocí operátoru „NEAR“).
- Umožňuje definování vnořených podmínek.
- Spolu s výsledky hledání zobrazuje i náhledy na dokumenty.
- Dovoluje hledání dle
 - geografické lokality
 - jazyka
 - typu souborů.
- Podporuje fonetické zadání dotazu.
- Umožňuje vyhledávání ve výsledcích.
- Prohledává audio, video a RSS soubory (v budoucnu nejen jejich meta data)
- Podporuje “regulární výrazy”.

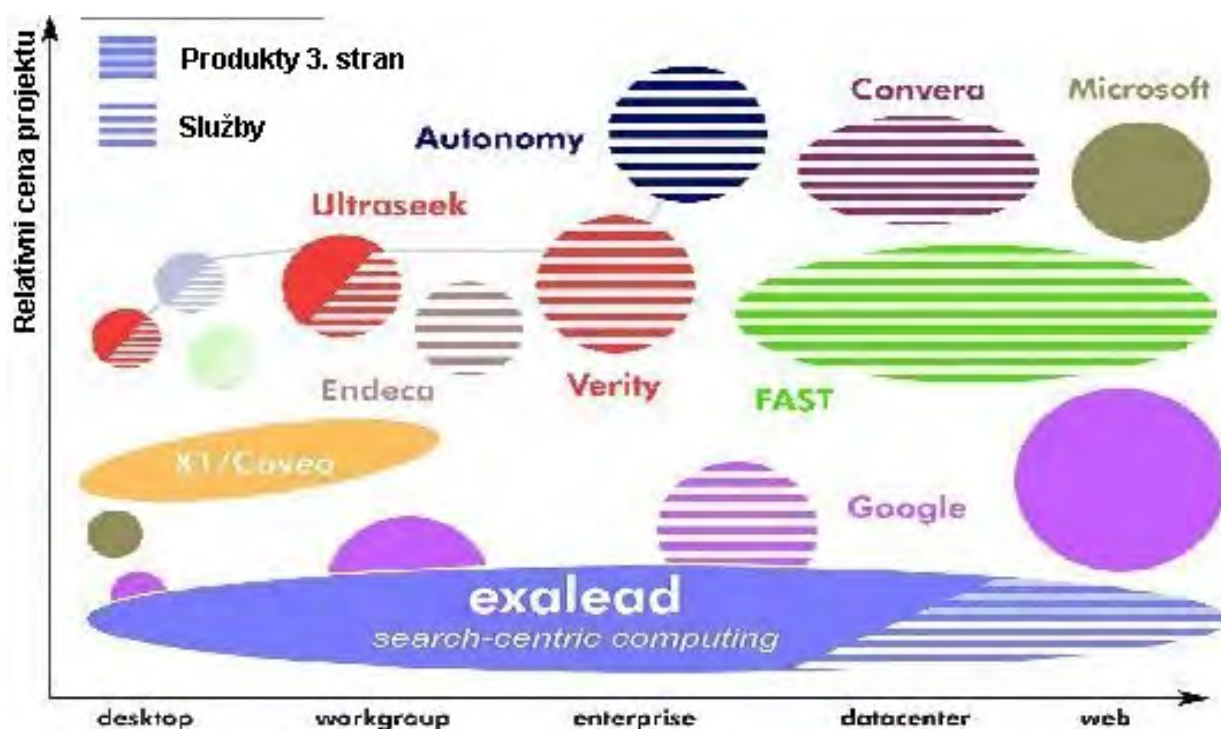
4.3 Základní schéma architektury

Na následujícím obrázku je základní schéma, ze kterého je patrné, ve kterém místě proudí informace se produkty pro vytěžování informací nachází. Systém indexuje definovaná úložiště a funkcionalitu pro vyhledávání zpřístupňuje přes standardní rozhraní Windows, Web a celou řadu dalších mobilních a specializovaných rozhraní.



4.4 Situace na trhu

Následující obrázek ukazuje, které produkty pro vyhledávání či správu informací jsou v současné době na trhu. Jak je patrné z obrázku, někteří dodavatelé řeší vytěžování informací pomocí několika produktů zaměřených na vyhledávání ve speciálních oblastech, někteří se snaží pokrýt potřeby více sektorů jedním produktem.



4.5 Reprezentativní systémy pro vyhledávání dat uvnitř organizací

V této části jsou uvedeny reprezentativní moduly, které slouží pro vyhledávání dat uvnitř organizací a na sdílených úložištích. Zároveň však jsou součástí rozsáhlejších systémů, které dovolují integrovat vyhledávání na vlastním osobním počítači, sdílených serverech, v databázích a na Internetu.

Základní verze, umožňující hledání na vlastním osobním počítači, bývají dostupné jako „freeware“, a proto jejich užití není spojeno s žádnými přímými licenčními poplatky. Teprve jejich napojení na servery společnosti vyžaduje další software, a tedy nákup licencí. Již nasazení těchto „volných“ verzí však výrazně usnadní práci a dovoluje integrovat hledání v osobním počítači a na Internetu.

Nasazení minimálních verzí vyhledávacích nástrojů je levné, a přesto ihned přináší výrazné zefektivnění práce. Zároveň otevírá možnost přechodu na systematická řešení vytěžování dat organizací.

4.6 Google Desktop

Odkaz pro stažení volné verze tohoto modulu naleznete na : <http://desktop.google.com/>

Mnoho informací je uloženo na disku ve vlastním počítači. Často tak, že již o těchto údajích ani nevíme a sháníme je tak složitě znovu v externích datových zdrojích.

Aplikace Google Desktop slouží právě k vyhledávání takových informací, ať už se nacházejí v poště, textových souborech, dokumentech kancelářských aplikací či navštívených stránkách Internetu. Podstatou je proces zvaný „indexování“, kdy se informace o obsahu dokumentů uchovávají v pracovních souborech systému tak, aby pozdější vyhledávání dokumentů podle části jejich obsahu proběhlo co nejrychleji.

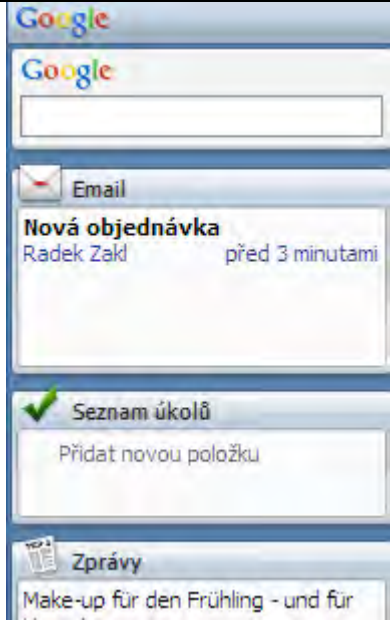
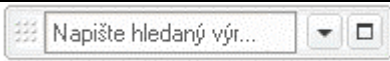
Indexování obsahu Vašeho PC se spouští automaticky po 30 ti sekundách nečinnosti. Dokumenty, se kterými právě pracujete, se indexují na pozadí Vaší práce a změny jsou tak zaznamenávány průběžně. První indexování může trvat i několik hodin, ale vzhledem k tomu, že probíhá v době, kdy s počítačem nepracujete, není tento proces na obtíž. Samozřejmě můžete procesu napomoci tím, že ponecháte systém zapnutý po skončení práce a proces indexace necháte běžet bez přerušení. Pomocí předvoleb je možné ovlivnit, které složky mají být indexovány a které ne.

4.6.1 Základy používání aplikace Google Desktop

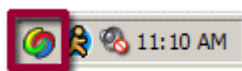
Vlastní proces hledání je totožný s tím, který znáte z běžného vyhledávače Google na Internetu. Po zadání kritérií kliknete na tlačítko **Prohledat počítač**.

Na vyhledání se dostanete jedním z následujících způsobů:

- Zadejte vyhledávaný termín do postranního panelu, lišty na hlavním panelu nebo do plovoucí lišty podle toho, jaké rozhraní Google Desktop je zvoleno (kliknutím pravým tlačítkem na ikonu Google Desktop na hlavním panelu vpravo dole můžete vybrat způsob zobrazení).

Postranní panel	Lišta na panelu (Deskbar)	Plovoucí lišta
		

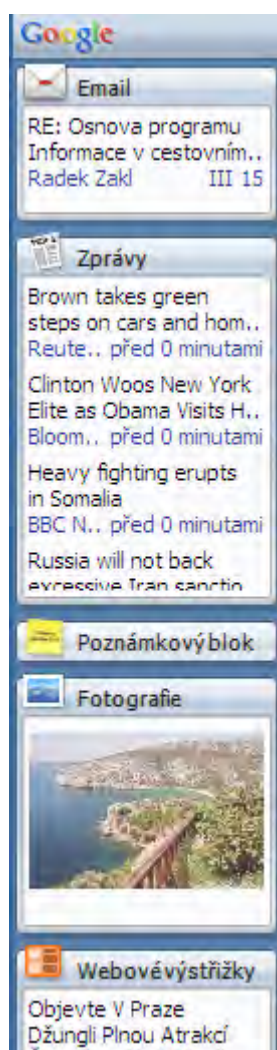
- Dvojitým klepnutím na ikonu Google Desktop na hlavním panelu.



- Kliknutím na odkaz Desktop nad vyhledávacím políčkem na domovské stránce Google.



Postranní lišta a „gadgets“



Google Desktop nabízí také panel, na kterém mohou být zobrazeny pro uživatele významné informace. Ty mohou pocházet z nejrůznějších vnějších i vnitřních zdrojů. Typicky budou pocházet z pošty a RSS kanálů, ale mohou sloužit také jako poznámkový blok či správce úkolů. Tyto jednotlivé prvky jsou seskupeny do rámečků zvaných gadgets (udělátka).

- Klepnutím na tlačítko se šipkou v nabídce modulu Google Gadget a volbou **Možnosti** lze nastavit chování těchto prvků dle Vašich představ.
- Táhnutím modulů Google Gadgets za titulní lištu je můžete na postranním panelu přeskupit nebo je od postranního panelu odpojit a umístit na plochu.
- Dvojitým klepnutím na záhlaví okna můžete moduly Google Gadgets sbalit nebo rozbalit.

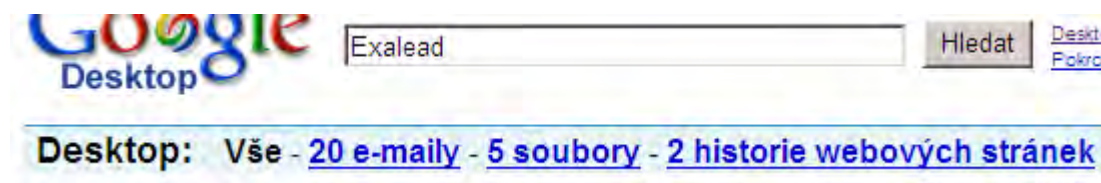
Na obrázku vlevo vidíte, jak vypadá postranní pruh s některými „gadgets“.

Na Internetu naleznete mnoho těchto „udělátek“, které Vám pomohou lépe organizovat Vaši práci.

4.6.2 Výsledky hledání

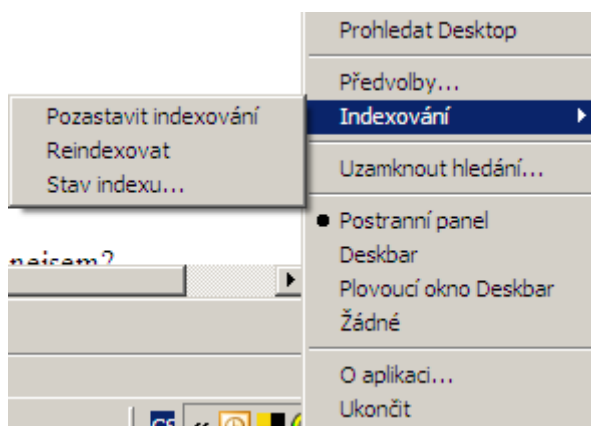
Poté co proběhne vyhledání, je zobrazen seznam položek, které odpovídají hledaným termínům. Jednotlivé položky volíte kliknutím na ně. Položky lze samozřejmě různě řadit.

V horní části obrazovky s výsledky můžete kliknutím na název příslušné skupiny souborů zobrazit seznam omezený jen na příslušný typ. Například kliknutím na výraz **20 – e-maily** se zobrazí seznam 20 ti nalezených položek pošty a ostatní typy souborů budou skryty.



Stačí klepnout na odkaz **Archiv** na stránce výsledků hledání a přejdete na verze dané položky, které má aplikace Google Desktop ve své vyrovnávací paměti. To je zvlášť užitečné, pokud nechtěně vymažete nějaký soubor!

4.6.3 Indexování a správa aplikace



Klikněte pravým tlačítkem myši na ikonu aplikace Google Desktop na hlavním panelu. Zde máte možnost pozastavit indexování, znovu vybudovat index a sledovat jeho stav. Při pozastavení indexování se proces zastaví na 15 minut. Zde je třeba připomenout, že se jedná o indexování souborů se kterými právě pracujete. Indexování ostatních souborů probíhá v době nečinnosti počítače.

Z tohoto menu lze také aplikaci Google desktop ukončit (po startu počítače se pak opět automaticky spustí). Aplikace se dá spustit též ze standardní nabídky **Programy → Google Desktop → Google Desktop**.

V této nabídce též naleznete možnost uzamknout hledání (když nechcete, aby ve Vašem počítači někdo prohledával soubory, když u něho právě nejste).

4.6.4 Pokročilejší hledání

V praxi samozřejmě potřebujete vyhledávat údaje na základě více podmínek, než je jen prostý výraz. Syntaxe výrazů již byla ukázána v části věnované vyhledávání na internetu, připomeneme si jen základní pravidla:

Poznámka: Výrazy jsou v tomto textu pro názornost označeny hranatými závorkami – ve vlastním okně vyhledávače se nepíší!

Fráze

Chcete – li najít přesnou frázi, uzavřete výraz do uvozovek.

Například: [„hotely ve Francii“]

Přítomnost více výrazů v jednom dokumentu

Operátor „+“

Chcete –li nalézt dokumenty , které obsahují všechny Vámi požadované výrazy, pak před ně napište operátor +.

Například výraz: [+Praha +koncert +filharmonie] nalezne všechny dokumenty ve Vašem počítači, které obsahují současně tyto tři výrazy.

Alternativou příkazu, který hledá více výrazů je uzavřít je do oblé závorky – [(Praha koncert)]

Poznámka: jen ty oblé závorky, ty hranaté jsou jen v tomto dokumentu, aby bylo jasné, co je výraz ☺.

Vyloučení výrazů

Operátor „-“

Chcete –li nalézt dokumenty , které obsahují některé požadované výrazy a neobsahují jiné pak použijte operátor „-“.

Například výraz: [+Praha +koncert -rock] nalezne všechny dokumenty ve Vašem počítači, které **obsahují** výrazy „Praha“ a „koncert“ a **neobsahují** výraz „rock“.

Prohledávání určitých stránek

Operátor "site:"

Tento operátor omezí výběr na danou Internetovou stránku.

Například výraz: [koncert site:www.praha.cz] zobrazí pouze stránky, které jste si prohlíželi na stránce www.praha.cz a obsahují slovo „koncert“.

Vyhledávání v určitých typech souborů

Operátor **"filetype:"**

Například výraz, který má za operátorem **filetype** příponu či název aplikace MS Office:

[obrat filetype:xls] nebo [obrat filetype:excel] slouží k vyhledání dokumentů Excelu, které obsahují slovo „obrat“.

Určení složek, kde se vyhledává

Operátor **"under:"**

Tento operátor dovoluje určit složky, ve kterých se hledaný soubor má nacházet.

Například výraz: [ubytování under:"C:\Partneři"] zobrazí pouze dokumenty obsahující výraz „ubytování“ ze složky **Partneři** na Vašem disku **C:**.

Cílené hledání e-mailů

Pro vyhledávání v poště jsou k dispozici následující operátory pro vyhledávání v příslušných sekcích zpráv:

Subject: (Předmět)

To: (Komu)

From: (Od)

Cc: (Kopie)

Bcc: (Skrytá kopie)

Například:

- Výraz [subject:hotel] vyhledá všechny e-maily se slovem "hotel" v poli Předmět:.
- Výraz [subject:(hotel Plzeň)] vyhledá všechny e-maily s oběma slovy "hotel" a "Plzeň" v poli Předmět:.
- Výraz [subject:hotel Plzeň] vyhledá všechny e-maily se slovem "hotel" v poli Předmět: a slovem "Plzeň" kdekoli v e-mailu, včetně těla zprávy.

V rámci jednoho vyhledání lze kombinovat omezení na jednotlivé sekce:

Například výraz [subject:hotel from:Radek] umožní vyhledat všechny zprávy s předmětem obsahujícím slovo „hotel“ od uživatele „Radek“.

Komplexní příklad na závěr:

Výraz [„úžasná Paris“ -Hilton from:Soňa] vyhledává mail od dámy jménem Soňa, ve kterém se píše o krásách Paříže a ne o krásách slečny Paris Hilton.

4.6.5 Řešení pro organizace

Výše popsané řešení umožňuje hledat data ve vlastním počítači (a Internetu). Mnoho informací je však uloženo na sdílených úložištích uvnitř organizace i mimo ni. Na takové vyhledávání je již třeba doplnit další nástroje. Společnost Gogole zde nabízí dva produkty:

- Google Mini pro menší organizace
- Google Search Appliance - pro střední a velké společnosti

Tyto systémy umožňují vytvářet sdílené indexy a vyhledávat data ve společných úložištích, ať už se jedná o interní (souborové servery) či externí zdroje (Internet). Typicky pracují za vnitřními hranicemi firewallu společnosti a indexují vlastní webové a souborové servery. Rozpoznávají více než 220 formátů a pracují se 190 ti jazyky.

Z technického hlediska se jedná o speciální servery, které se instalují formou „plug and play“. Mají vlastní operační systém a zákazník je obdrží již nainstalované. Pouze se začlení do síťové infrastruktury a konfiguruje se.



Na obrázku je vidět, že tyto servery se mohou seskupovat a pokrýt tak potřebu vyhledávání i v extrémně náročném prostředí.

Z pohledu koncového uživatele je ovládání stejné, jako to, které jsme ukázali na aplikaci Gogole Desktop.

4.6.6 Další možnosti rozšíření

V základní podobě vyhledává Gogole Desktop v následující sadě souborů:

Email (Outlook, Gmail), Chat, Historie Webu, Soubory médií, Textové soubory, Soubory MS Office, Akrobat PDF, Zip, Kontakty, Úkoly, Kalendář, poznámky.

Existuje však možnost napojení na celou řadu dalších systémů, a to pomocí tak zvaných Plug-Ins (<http://desktop.google.com/plugins/>). V sekci věnované indexování na uvedeném odkazu naleznete seznam takovýchto dostupných modulů.

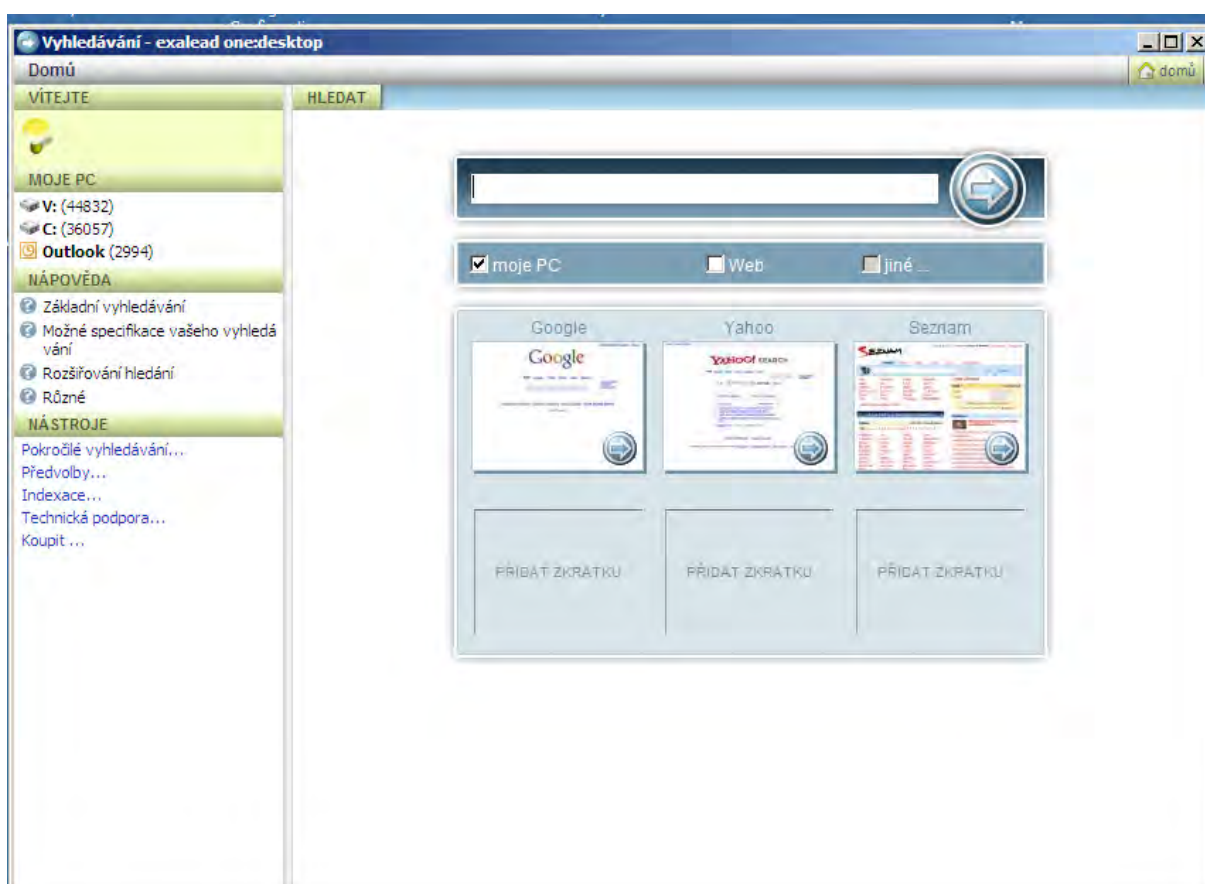
4.6.7 exalead one:desktop

Odkaz ke stažení volné verze tohoto modulu naleznete na: <http://corporate.exalead.com>

Dalším dodavatelem, který se soustředí na aplikace pro vyhledávání a organizaci dat je francouzská společnost Exalead. Zde se budeme zabývat aplikací **Exalead one:desktop**. Její místo v infrastruktuře společnosti pro vyhledávání je stejné, jako tomu je v případě aplikace Google Desktop. I tato aplikace vytváří index pro snadné vyhledávání informací ve vlastním počítači, ať už se tyto informace nacházejí v poště, textových souborech, dokumentech kancelářských aplikací či na navštívených stránkách Internetu.

4.6.8 Základy používání aplikace exalead one:desktop

Následující snímek ukazuje pracovní obrazovku modulu:



Vždy jsou viditelné určité sekce v levém sloupci, které při spuštění obsahují Nápovědu a Nástroje pro předvolby a nastavení. V průběhu práce se obsah sloupce mění a nabízí potřebnou funkcionalitu. Hlavní sekce pro zadání podmínek pro vyhledávání je při spuštění uprostřed.

Zde je možné volit, zda chcete hledat informace pouze ve Vašem osobním počítači (moje PC) nebo i v Internetu (Web) či jiném datovém úložišti.

Panel také obsahuje odkazy - náhledy na vybrané internetové portály, které můžete volit dle vlastních preferencí.

4.6.9 Výrazy pro hledání

Poznámka: Výrazy jsou v tomto textu pro názornost označeny hranatými závorkami – ve vlastním okně vyhledávače se nepíší!

Základní vyhledávání

Zdáte – li například výraz [hotel Praha], pak systém nalezne dokumenty obsahující oba výrazy nebo dokumenty obsahující jen některý z nich. Ty soubory, které obsahují oba výrazy budou uvedeny ve výsledném seznamu nejdříve.

Rozšířené (pokročilé) vyhledávání

Použijete – li „pokročilejší volby“, můžete vyhledávat „chytřejšími“ metodami, například:

„Slovní Přizpůsobení“ či „Automatický slovní kmen“

Při tomto nastavení se do výsledku zahrnou i jiná skloňování a nerozlišuje se množné a jednotné číslo.

„Fonetické vyhledávání“

Dovoluje vyhledávání i podle podobnosti slov. Zde nejlépe poslouží příklad:

Pomocí gramaticky chybného výrazu [Fylmová hvězda] se najdou také dokumenty, kde je správně uvedeno „filmová hvězda“.

Přesné fráze

Jestliže chcete vyhledat dokumenty, které obsahují přesné fráze, pak uzavřete příslušnou frázi do uvozovek.

Například : ["rodinný hotel"] vyhledá dokumenty, kde se nachází tato přesná fráze.

„Přibližné“ vyhledávání

Pomocí operátoru **NEAR** se vyhledávají soubory (dokumenty), kde jsou vyhledávané termíny ve vzdálenosti maximálně 16 slov od sebe.

Obecně lze zadat **NEAR/x** kde x je počet slov mezi výrazy.

Operátor **NEXT** najde dokumenty kde jsou hledané termíny hned za sebou.

Například : [filmová hvězda NEAR Hollywood] najde dokumenty, kde se výraz **filmová hvězda** vyskytuje v těsné blízkosti slova Hollywood.

[filmová NEXT hvězda] prohledá dokumenty, kde se slova **filmová** a **hvězda** objevují hned za sebou v daném pořadí.

Přesná slova

Operátor **+(plus)**

Tento operátor se hodí, chcete - li, aby byly vyhledány pouze dokumenty obsahující přesné znění slova - (nepoužije se volba „**Slovní přizpůsobení**“). Také umožňuje hledat spojky a předložky (a, i, od, nebo) které se ve výchozím nastavení ignorují.

Například : výraz [+Hotely] vyhledá dokumenty obsahující přesně slovo **Hotely** nenajde se např. **Hotel**.

Specifikace cizího jazyka

Operátor **language:**

Použitím **language:XX** (XX = FR, GE, atd.) se omezí hledání na dokumenty psané v daném jazyce.

Vyhledávání na určené webové adrese.

Operátor **site:**

Tento operátor umožňuje vyhledávat informace jen v určité doméně (na určitých webových stránkách).

Například: výraz [luxusní hotely site:www.praha.cz]

najde na adrese **www.praha.cz** dokumenty obsahující slova **luxusní** a/nebo **hotely**.

Hledání v titulku dokumentu

Operátor **intitle:**

Pomocí tohoto operátoru naleznete dokumenty, které obsahují hledané výrazy v titulku dokumentu.

Například: [intitle:"all inclusive"] najde dokumenty, které mají výraz „**all inclusive**“ v titulku.

Vylučovací operátor

Operátor **-(mínus)**

Pomocí tohoto operátoru můžete vyloučit z výběru dokumenty obsahující výrazy, které v dokumentu nechcete.

Například: [hotel -exclusive] najde dokumenty obsahující slovo **hotel** a neobsahující slovo **exklusive**.

Vyhledávání podle data

Operátor **date**

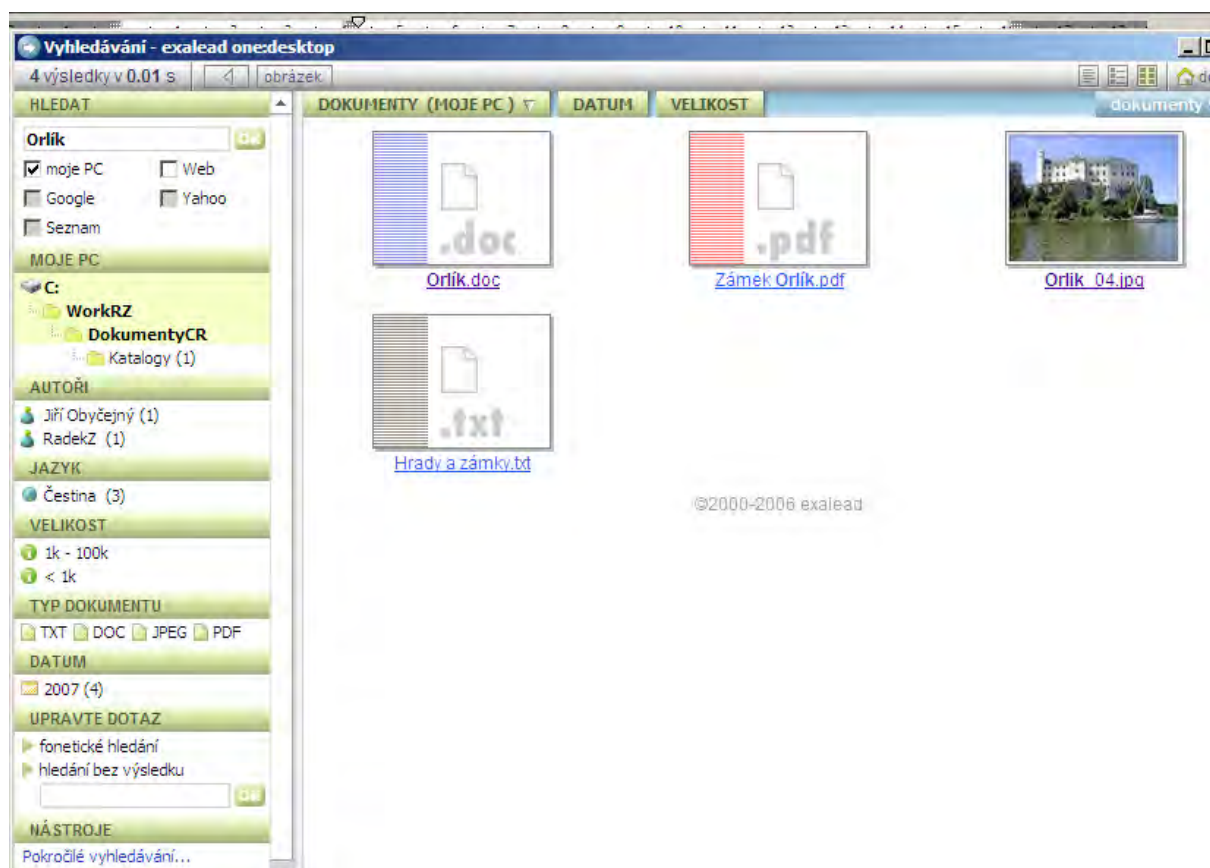
Pomocí tohoto operátoru vyhledáváme dokumenty mladší či starší určitého data. Jde přitom o datum vytvoření či poslední modifikace souboru.

Například: [„hotel Astra“ date>=2006/12/07] nalezne dokumenty obsahující frázi „hotel Astra“ a vytvořené či modifikované 7.prosince 2006 a později.

4.6.10 Uspořádání výsledku

Od inteligentních vyhledávacích systémů očekáváme samozřejmě mnohem víc, než je jen nalezení požadované informace. Jednou z důležitých vlastností je třídění a uspořádání nalezených dokumentů dle určitých kritérií, tak, aby se v nich koncový uživatel rychle orientoval.

Příklad je patrný na následujícím obrázku:



Jedná se o výsledek hledání dokumentů obsahujících slovo „**Orlík**“. V hlavní části obrazovky jsou vidět jednotlivé nalezené soubory. Podle náhledu je zřejmé, o jaký dokument se jedná.

V levé části je vidět roztrídění výsledků dle určitých kritérií:

Složka	Obsah
MOJE PC	Zde vidíme, kde na disku se soubory nacházejí. Závorka značí, kolik dokumentů ve které složce splňuje výběrová kritéria.
AUTORI	Kliknutím na autora se zobrazí jen ty soubory, které vytvořil.
JAZYK	Zde jsou všechny dokumenty v češtině.
VELIKOST	Soubory se třídí též do skupin dle velikosti.
TYP DOKUMENTU	Zde vidíme typy souborů, které se našly. Opět kliknutím na vybraný typ se zobrazí pouze soubory vybraného typu.
DATUM	Kliknutím na 2007 se rozbalí hierarchie dní, ve kterých dokumenty vznikly.

4.6.11 Hledání ve výsledku

Výsledkem hledání může být velké množství souborů a dobrý systém by měl umožnit hledat i uvnitř sady nalezených dokumentů.

Exalead zde nabízí volby:

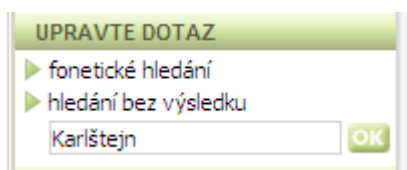
Fonetické hledání

Po stisknutí tohoto odkazu se dotaz zopakuje, ale použije se funkce fonetického hledání (dovoluje vyhledávání i podle podobnosti slov či výslovnosti slov).

To může být velmi účelné, hledáte – li dokumenty obsahující například jména cizinců či cizí lokality, u nichž si nejste úplně jisti tím, jak se píší.

Hledání ve výsledku

Poznámka: v současné verzi je chybně v produktu přeloženo „Search within results“ jako „Hledání bez výsledku“, ale myslím, že se tento překlep dá obejít s úsměvem.



Zadáte – li do pole **hledání ve výsledku** další podmínku, jako třeba na výše uvedeném obrázku „Karlštejn“, pak se v sadě nalezených dokumentů hledají takové, které toto slovo obsahují.

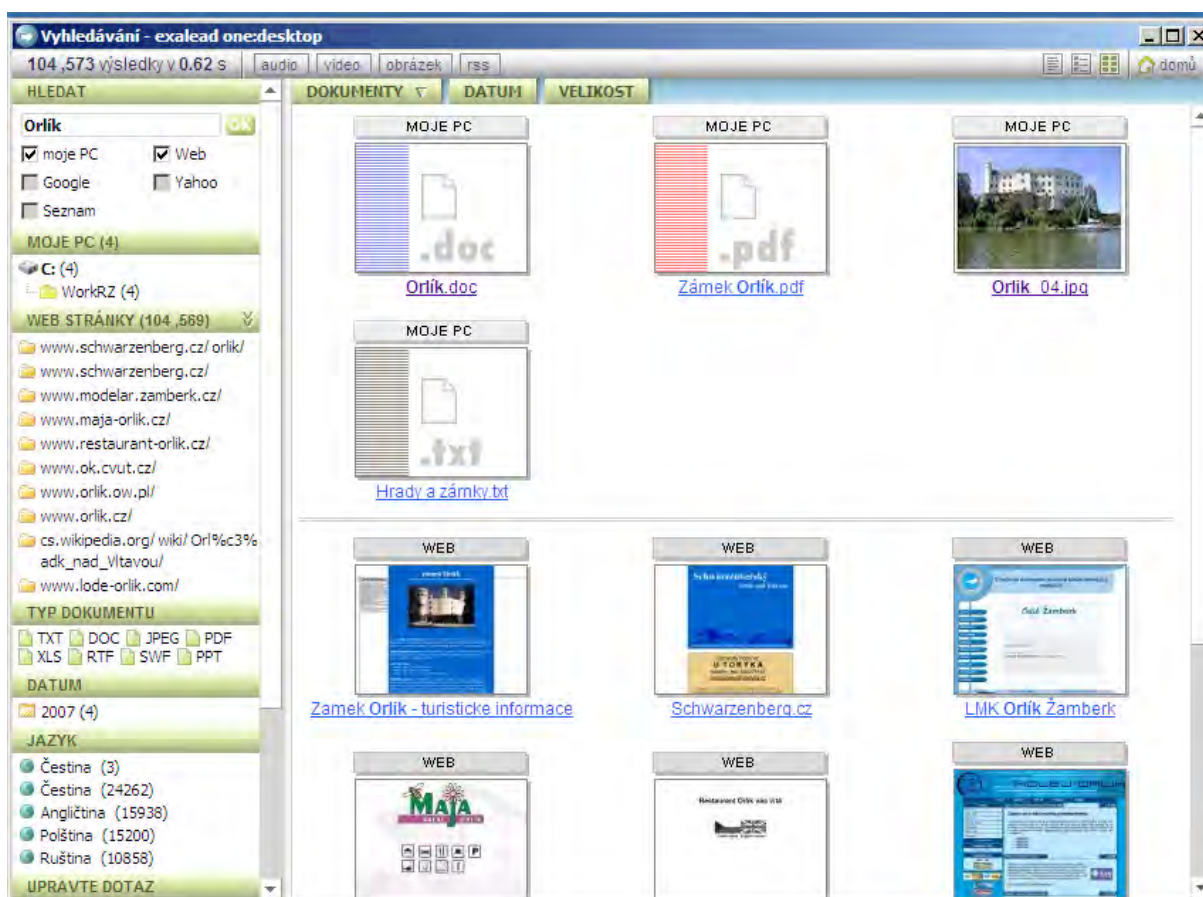
Výsledkem je pak vlastně kombinace (průnik) podmínek, v našem případě dokumenty obsahující slova „Orlík“ a „Karlštejn“.

4.6.12 Vyhledávání ve vlastním PC a na Internetu současně

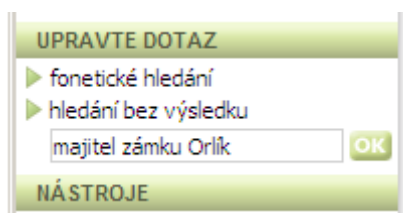
Velkou silou vyhledávacích systémů je propojení více datových zdrojů. Už v nejjednodušší podobě produktů **Gogole Desktop** i **Exalead one:desktop** je možné vyhledávat současně ve vlastním PC a na Internetu.

K aktivaci postačí zaškrtnout v horní levé části v sekci **HLEDAT** políčka **moje PC** a **Web**. Ve výsledku se zobrazí jak dokumenty nalezené v osobním počítači, tak v Internetu. K vyhledání na Internetu se použije ve standardní instalaci vyhledávač Exalead, nicméně lze též nastavit vyhledávání jakýmkoli oblíbeným vyhledávačem, popřípadě je kombinovat.

Dejme tedy vyhledat ve vlastním počítači (moje PC) i na Internetu (Web) vše, co obsahuje slovo **Orlík**.



Ve výsledku je vidět, že informace vskutku pocházejí z mnoha typů souborů a mnoha jazykových prostředí. Je zřejmé, že prohledávat všechny soubory, abychom našli řekněme **majitele** zámku Orlík, by bylo zdlouhavé. Nejprve tedy klikneme na odkaz **Čeština**, abychom se nadále zabývali jen dokumenty psanými česky a poté do okna úpravy dotazu (viz. obrázek)



napíšeme výraz **majitel zámku Orlík**. Po stisknutí odkazu OK se na prvním místě ukáže stránka odkazující na historii zámku a v dolní části náhled této stránky.



Již z náhledu vidíme, že to bude to pravé místo, kde se minulých i současných majitelích zámku dozvíme vše co potřebujeme.

4.6.13 Řešení pro organizace

Tak, jako tomu bylo u společnosti Google, tak i Exalead nabízí řešení pro vyhledávání ve sdílených úložištích organizací (připomeňme, že předchozí sekce dokumentu hovořila o vyhledávání ve vlastním PC a/nebo v Internetu).

Exalead nabízí řadu produktů, odstupňovanou dle množství údajů, které indexuje (a ve kterých umí vyhledávat). Na rozdíl od společnosti Google se jedná o řešení softwarové (součástí dodávky není žádný hardware). Stejně tak, jako u společnosti Google však licence software pro tato řešení již nejsou zdarma.



- exalead one:Workgroups
 - je určen pro středně velké společnosti nebo samostatná oddělení větších organizací
- exalead one:enterprise
 - pro velké organizace
- exalead one:datacenter
 - pro organizace produkující extrémní množství nestrukturovaných dat

4.7 Použitá literatura

1. EXALEAD, *Exalead - Products - exalead one:desktop* [online]. c.2007, last revision 22 March 2003 [cit. 2007-03-22] Dostupné z <<http://corporate.exalead.com>>
2. GOOGLE, *Google Desktop - Funkce* [online]. c.2007, last revision 22 March 2007 [cit. 2007-03-22] Dostupné z <<http://desktop.google.com/cs/features.html>>
3. GOOGLE, *Google Desktop - Firemní verze* [online]. c.2007, last revision 22 March 2007 [cit. 2007-03-22] Dostupné z <<http://desktop.google.com/cs/enterprise/index.html>>
4. GOOGLE, *Google Desktop – Help Center* [online]. c.2007, last revision 22 March 2007 [cit. 2007-03-22] Dostupné z <<http://desktop.google.com/support/?hl=cs>>

REJSTŘÍK POJMŮ

- arpanet, 19
- boti, 26
- CERN, 19
- crawleři, 26
- description, 47
- desktop, 61, 65, 68, 69, 75
- DHCP, 20
- Directory of MOZilla, 24
- DNS, 20
- e-mail**, 56, 66
- ERP, 56, 57
- Exalead**, 69, 74, 75, 77
- Fonetické vyhledávání**, 70
- Fráze**, 65
- freeware, 61
- gadgets, 63
- Gogole Desktop, 61, 68, 75
- Google, 56, 61, 62, 63, 64, 65, 67, 69, 77
- Google bomba, 49
- Google Desktop, 62, 65
- Google, Inc., 28
- Hluboký, neviditelný web, 23
- HTTP, 20
- Hyper Text Transfer Protocol, 20
- Hypertext, 21
- indexování, 61, 64, 68
- Indexování, 61, 64
- Internet, 18, 56, 58, 61, 62, 67, 69, 70, 75
- katalog, 39
- NFS, 20
- Open Document Standard, 22
- Outlook, 56, 57, 68
- PageRank, 25, 26
- PDF, 21
- Prostý text, 21
- Protokol, 20
- Rich Text Format, 21
- Search Engine Optimization, 47
- SEO, 47
- SMTP, 19
- SNMP, 20
- sponzorovaný odkaz, 33
- S-Rank, 25
- Telnet, 19
- vyhledávací fráze, 32
- vyhledávač, 24
- vyhledávání obrázků, 41
- Vyloučení výrazů**, 65
- Windows, 56, 59
- Windows Explorer, 56
- World Wide Web, 21

Tato skripta jsou spolufinancována
Evropským sociálním fondem a státním rozpočtem České republiky