

Internet

Internet poskytuje svým uživatel celou řadu nejrůznějších služeb.

http(S)

Http neboli **Hypertext Transfer Protocol** je protokol pro přenos dokumentů v Internetu, používá se nejčastěji na WWW. 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.

🌐 O cookies více v RFC 2965 (<http://www.ietf.org/rfc/rfc2965.txt>)

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

🌐 Webové prohlížeče:

- Apple Safari (<http://www.apple.com/safari/>)
- Google Chrome (<http://www.google.com/chrome>)
- Internet Explorer (<http://www.microsoft.com/windows/internet-explorer/>)
- Mozilla Firefox (<http://www.mozilla.com/firefox/>)
- Opera (<http://www.opera.com/>)

E-mail

Další službou je elektronická verze dopisu zvaná **e-mail**, neboli **elektronická pošta**. Ta funguje na principu výměny zpráv mezi počítači pomocí Simple Mail Transfer Protocol (SMTP) a softwaru typu MTA (mail transfer agent neboli mail transport agent, message transfer agent, mail server nebo mail exchange) je počítačový program, který přenáší elektronickou poštu z jednoho počítače na druhý.

K práci s e-mailovým účtem může většinou posloužit webové rozhraní, nicméně k plnohodnotné práci s e-maily je vhodné použít program – poštovní (e-mailový) klient. Ten stahuje zprávy z poštovního serveru použitím buď protokolů POP nebo IMAP nebo některého komerčního protokolu jako např. Lotus Notes nebo Microsoft Exchange Server.

Elektronická pošta využívá pro své fungování několik protokolů. SMTP je protokol zajišťující doručení (odeslání) pošty pomocí přímého spojení mezi odesílatelem a adresátem. Protokoly POP a IMAP zajišťují příjem (stažení) emailových zpráv. Tyto protokoly používají v současné době téměř všichni uživatelé elektronické pošty.

IMAP (Internet Message Access Protocol) vyžaduje trvalé připojení (tzv. on-line), nabízí pokročilé možnosti vzdálené správy (práce se složkami, přesouvání zpráv, prohledávání na straně serveru apod.). Díky tomu je možné s celou poštovní schránkou plně pracovat z libovolného místa. Všechny zprávy a složky jsou uloženy na poštovním serveru a na počítač se stahují jen nezbytné informace, takže při zobrazení složky se stáhnou jen záhlaví zpráv a jejich obsah až v případě, že zprávu chce uživatel přečíst. U jednotlivých zpráv se uchovává jejich stav (nepřečtená, zodpovězená, důležitá), uživatel může zprávy přesouvat mezi složkami, složky vytvářet, mazat, prohledávat na straně serveru apod. Protokol umožňuje současné připojení více klientů zároveň. Protokol IMAP je tedy zvláště vhodné využít, pokud přistupujeme ke stejnému účtu současně z práce, z domova a dalších míst a nevystačí nám webové rozhraní e-mailu. Aktuální verze protokolu je IMAP4.

POP (Post Office Protocol) je internetový protokol, který se používá pro stahování emailových zpráv ze vzdáleného serveru na klienta. Na rozdíl od IMAPu pracuje POP jednostranně. Ze vzdáleného serveru se stáhnou všechny zprávy, včetně těch nechtěných (spam apod.). Tuto nevýhodu může odstranit protokol IMAP, který pracuje se zprávami přímo na serveru. Dnes se již používá 3. verze těchto protokolů, nazývaná POP3.

Protokol	Port pro nezabezpečenou komunikaci	Port pro zabezpečenou komunikaci
SMTP	25	465
IMAP4	143	993
POP3	110	995

🌟 E-mailový klienti:

- Eudora (<http://www.eudora.com/>)
- Lotus Notes (<http://www-01.ibm.com/software/lotus/notesanddomino/>)
- Outlook (<http://office.microsoft.com/outlook>)
- The Bat! (<http://www.ritlabs.com/en/products/thebat/>)
- Thunderbird (<http://www.mozillamessaging.com/thunderbird/>)
- Zimbra (<http://www.zimbra.com/>)

FTP

FTP (File Transfer Protocol) je protokol určený pro přenos souborů mezi počítači. Protokol je interaktivní a umožňuje řízení přístupu (přihlašování login/heslo), specifikaci formátu přenášeného souboru (znakově - binárně), výpis vzdáleného adresáře atd. Nejčastěji se FTP používá pro sdílení dat (často hudba, videa, vlastní tvorba, apod.) nebo správu účtů internetových stránek. FTP zpřístupňuje systémy souborů na jiném počítači a to nejčastěji prostřednictvím programů na používání FTP, u kterých si nemusíme pamatovat strukturu příkazů a pracují velice podobně jako běžný průzkumník (např. oblíbený Total Commander také podporuje FTP).

Při běžném připojování pomocí protokolu FTP jsou přihlašovací údaje (jméno a heslo) přenášeny v textové podobě a je technicky možné je odchytnout. Následně mohou být data zcizena nebo upravena. Často se vyskytují zprávy o tom, jak se hackeři nabourali do serverů a nezdá se jim k tomu stačil tak triviální úkon, jakým je zachycení nešifrovaného FTP hesla. Netřeba zdůrazňovat, že následky těchto útoků mohou mít katastrofální dopad. V dnešní době je dostupná zabezpečená verze tzv. FTPS, která je zpětně kompatibilní s běžným FTP.

🏠 O FTP protokolu více v RFC 959 (<http://www.ietf.org/rfc/rfc959.txt>)

🏠 Klienti pro práci s FTP:

- FileZilla (<http://filezilla-project.org/>)
- FlashFXP (<http://www.flashfxp.com/>)
- Total Commander (<http://www.ghisler.com/>)
- WinSCP (<http://winscp.net/eng/docs/introduction>)

Instant messaging

Instant messaging (Tencent QQ, Windows Live Messenger, Yahoo! Messenger, Skype, AIM, Jabber, eBuddy, IBM Lotus Sametime, ICQ, Xfire a další) je internetová služba, umožňující svým uživatelům sledovat, kteří jejich přátelé jsou právě připojeni, a dle potřeby jim posílat zprávy, chatovat, přeposílat soubory mezi uživateli a i jinak komunikovat. Hlavní výhodou oproti používání např. e-mailu spočívá v principu odesílání a přijímání zpráv v reálném čase. Jinými slovy zpráva je doručena ve velmi krátké době od odeslání (většinou v rámci stovek milisekund).

Instant messaging zrychluje komunikaci a umožňuje snadnou spolupráci mezi více lidmi. Na rozdíl od e-mailu nebo telefonu druhá strana ví, zda je účastník k dispozici či nikoliv. Většina IM systémů umožňuje nastavit away message, tedy zprávu podle které lze zjistit, zda je uživatel přítomen přímo u svého počítače. Na druhou stranu uživatele nikdo nenutí, aby na zprávy odpovídali ihned. Tímto způsobem se IM komunikace stává méně vyrušující než třeba telefon a to je částečný důvod, proč je tento způsob komunikace stále více oblíben v obchodním prostředí. Instant messaging je ideální pro rychlou výměnu internetových adres, kusů zdrojového kódu a dalších věcí, které se např. v telefonní komunikaci špatně přenášejí.

🏠 Domovské stránky některých služeb:

- ICQ (<http://www.icq.com/>)
- Jabber (<http://www.jabber.org/>)
- Skype (<http://www.skype.com/>)
- Windows Live Messenger (<http://download.live.com/?sku=messenger>)

VoIP

Voice over Internet Protocol (zkratkou VoIP, také se používají termíny IP Telephony, Internet telephony, Broadband telephony, Broadband Phone a Voice over Broadband) je technologie, umožňující přenos digitalizovaného hlasu prostřednictvím počítačové sítě nebo jiného média, dostupného pro protokol IP. Využívá se pro telefonování prostřednictvím Internetu, intranetu nebo jakéhokoli jiného datového spojení.

Hlavní výhodou telefonování po Internetu jsou nižší náklady oproti klasické pevné lince, a často o desítky procent. Nutnou podmínkou pro kvalitní fungování je plnohodnotné internetové připojení. Není vhodný klasický Dial-Up ani GPRS (příliš pomalý), ani CDMA (moc dlouhá odezva). Doporučovaná minimální rychlost přenosu je 128 kb/s. Důraz je kladen hlavně na kvalitu připojení; odezvy by měly být krátké, nemělo by nedocházet k výpadkům.

Pro telefonování přes internet je třeba kromě zmíněného připojení telefonní přístroj (buď tzv. IP telefon, nebo klasický telefon + speciální převodník, nebo PC se zvukovou kartou a mikrofonom) a služby pro některého operátora.

🏠 další informace na <http://www.freevoice.cz/> nebo <http://www.telefonujeme.cz/>

VoIP operátoři:

- 802.VOX (<http://www.802.cz/>)
- viphone (<http://viphone.cz/>)
- FAYN (<http://www.fayn.cz/>)
- VOLNÝ TelefoNet (<http://volani.volny.cz/>)
- SIPY (<http://www.sipy.cz/>)

P2P

Síť peer-to-peer (doslova rovný s rovným), P2P nebo klient-klient je označení architektury počítačových sítí, ve které spolu komunikují přímo jednotliví klienti (uživatelé). Dnes se označení P2P vztahuje hlavně na výměnné síť, prostřednictvím kterých si mnoho uživatelů může vyměňovat data. Příkladem takových sítí jsou např. Gnutella, Torrent, Direct Connect, DC++, nebo eDonkey. Opakem je architektura klient-server, ve které jednotliví klienti komunikují vždy s centrálním serverem či servery, prostřednictvím kterého i komunikují s jinými klienty, pokud je to potřeba. Čistá P2P architektura vůbec pojem server nezná, všechny uzly sítě jsou si rovnocenné (a působí současně jako klienti i servery pro jiné klienty). V praxi se však často pro zjednodušení návrhu v protokolu objevují specializované servery, které ovšem slouží pouze pro počáteční navázání komunikace, „seznámení“ klientů navzájem, popř. jako proxy server v případě, že spolu z nějakého důvodu nemohou koncové uzly komunikovat přímo.

Jednou ze základních výhod P2P sítí je fakt, že s rostoucím množstvím uživatelů celková dostupná přenosová kapacita roste, zatímco u modelu klient-server se musí uživatelé dělit o konstantní kapacitu serveru, takže při nárůstu uživatelů klesá průměrná přenosová rychlost.

Nejčastějším obsahem šířeným po výměnných sítích jsou hudební nahrávky ve formátu MP3, filmy ve formátu MPEG a software. Velká část (pravděpodobně výrazná většina) tohoto obsahu je šířena bez souhlasu držitele autorských práv, tudíž v rozporu se zákonem. Důsledkem toho jsou žaloby na provozovatele takových sítí (hlavně v USA), podávané zástupci autorů a organizacemi jako RIAA či MPAA. Síť Napster byla kvůli úspěšné žalobě organizace RIAA zrušena (poté byl Napster odkoupen a pod stejnou značkou nyní nabízí legální prodej hudby přes Internet), další síť mnoha podobným žalobám čelí; zástupci autorských svazů často lobbují za přijetí tvrdších zákonů, které by výměnné síť postihovaly.

Dnešní anonymní výměnné síť umožňují (legální i nelegální) výměnu souborů s prakticky nulovou mírou odpovědnosti jednotlivých uživatelů. Někteří kritici poukazují na to, že se prostřednictvím výměnných sítí může distribuovat dětská pornografie či podporovat terorismus a na základě toho žádají o regulaci či přímo zákaz takových sítí. Zastánci naopak argumentují tím, že možnost zneužití technologie k nezákonným účelům nesmí bránit jejímu legálnímu využívání a že je třeba dodržovat princip presumpce nevinny.

RSS (Really Simple Syndication)

RSS je rodina XML formátů určených pro čtení novinek na webových stránkách a obecněji syndikaci obsahu.

Technologie RSS umožňuje uživatelům Internetu přihlásit se k odběru novinek z webu, který nabízí RSS zdroj (RSS feed, též RSS kanál, RSS channel). Tento zdroj se většinou vyskytuje na stránkách, kde se obsah mění a přidává velmi často (například zpravodajské servery).



RSS formát poskytuje obsah celého článku, příp. jeho část, odkaz na původní článek a také jiná metadata. Tyto informace jsou posílány jako XML soubor nazývaný RSS zdroj, webový zdroj, RSS stream, RSS feed nebo RSS kanál.

Software určený k práci s RSS kanály se označuje jako RSS čtečka. Může se jednat o samostatný specializovaný program, o plugin do jiného programu (typicky webového prohlížeče či programu pro instant messaging), tato funkce může být v jiném programu rovnou vestavěna, případně se může jednat o webovou aplikaci poskytující tuto funkčnost (např. Google Reader či NetVibes).

Pokud některý webový server nabízí RSS kanály, obvykle to indikuje ikonkou, která vede přímo na URL příslušného zdroje, který uživatel zadá do čtečky a ta poté zobrazí seznam všech takto zpřístupněných článků (či jiných odkazů). Čtečka pak pravidelně kontroluje toto URL a zobrazuje nové položky.

🏠 Vše podstatné o RSS (<http://www.root.cz/clanky/vse-podstatne-o-rss/>)

Atom (standard)

Atom (plným názvem Atom Syndication Format) je webový standard pro publikování syndikovaného obsahu. Je nástupcem formátu RSS. Kanál i jednotlivé položky musí mít uvedeny svůj název, jedinečný identifikátor (URI) a datum poslední změny. Povinné je rovněž jméno autora u každé položky, pokud není vyplněno pro všechny z nich. Zatímco u RSS bylo možné v poli description uvést jak souhrn, tak plný obsah, u Atomu je pro souhrn vyčleněn element summary, zatímco pro plný obsah se používá content. V RSS nebylo možné zvolit, jaký formát obsahu je použit. Atom rozeznává např. prostý text, escapované HTML, XHTML, XML, binární obsah v kódování Base64 či odkaz na jiný webový zdroj.

🏠 Specifikace RFC 4287 (<http://tools.ietf.org/rfc/rfc4287.txt>)

Systém webové adresy

TLD - Top-level domains

Doména nejvyššího řádu

Doména nejvyššího řádu (TLD, z anglického Top Level Domain) je internetová doména na nejvyšší úrovni stromu internetových domén. V doménovém jméně je doména nejvyšší úrovně uvedena na konci (např. u cs.wikipedia.org je doménou nejvyššího řádu org). TLD popisuje základní skupinu doménových jmen, např. všechna doménová jména daného státu. Domény nejvyššího řádu jsou pevně stanoveny internetovou standardizační organizací IANA. Existují TLD následujících tří typů:

1. **Národní** TLD (country-code TLD, ccTLD) sdružující domény jednoho státu. Jejich název je dvoupísmenný, až na výjimky odpovídající kódu země podle ISO 3166-1, např. cz pro Česko.
2. **Generické** TLD (generic TLD, gTLD) sdružující obecné domény (např. org pro neziskové organizace), nespojené s jedním konkrétním státem (až na výjimku TLD mil a gov, které jsou z historických důvodů vyhrazeny pro vojenské, resp. vládní počítačové sítě v USA).
3. **Infrastrukturní** TLD využívané pro vnitřní mechanismy Internetu. V současné době existuje jediná taková TLD: arpa, používaná systémem DNS.

V každém státu určuje příslušný správce národní domény pravidla, podle kterých se doménová jména v dané TLD přidělují.

🏠 CZ NIC – správce domén v ČR (<http://www.nic.cz/>)

Některé státy umožňují registraci jména v národní TLD libovolnému zájemci, bez ohledu na to, jestli má s příslušným státem něco společného. Mezi takové státy a území patří i Česko se svou doménou .cz, dále také např. Arménie (.am), Itálie (.it), Jersey (.je), Kokosové ostrovy (.cc), Německo (.de), Niue (.nu), Rakousko

(.at), Samoa (.ws), Tonga (.to), Turkmenistán (.tm) či Tuvalu (.tv). To umožňuje existenci „zajímavých“ doménových jmen jako např. coje.to, I.am (angl. jsem), start.at (angl. začít u), go.to (angl. jít na).

Jiné státy a závislá území povolují registraci domény v příslušné TLD pouze občanům či podnikům té které země (např. Kanada, Mongolsko či Slovensko).

🏠 Většina národních domén používá kód země podle ISO 3166-1 (http://www.iso.org/iso/country_codes/iso_3166_code_lists.htm)

Některé ccTLD však tuto konvenci nedodržují.

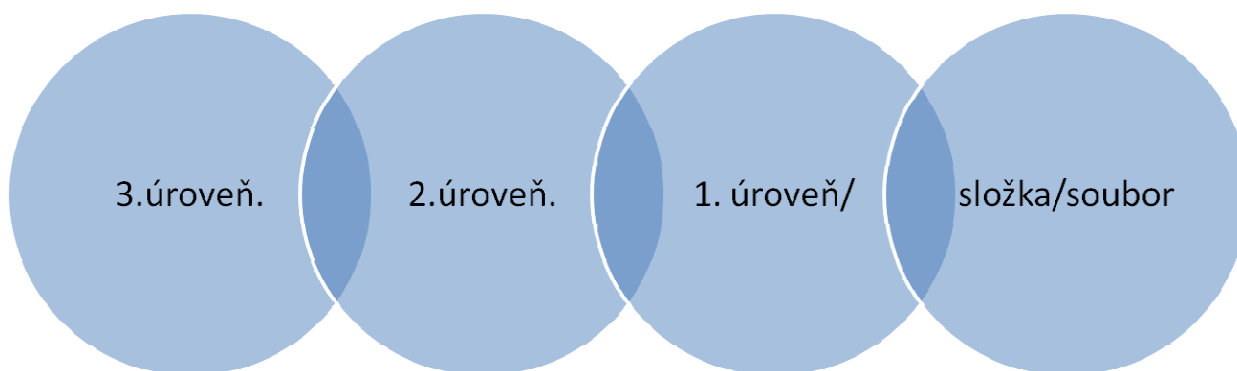
Generická doména nejvyššího řádu (anglicky generic top-level domain, zkráceně generic TLD nebo gTLD) je doména nejvyššího řádu (alespoň teoreticky) společná pro daný typ subjektů. Její název je nejméně třípísmenný.

Generické domény lze rozdělit na:

- Neomezené, jejichž použití není nijak omezeno (např. .com, .info, .net, .org, nebo .eu)
- Vyhrazené, které jsou určeny pro specifický účel, ale není to vyžadováno (např. .name)
- Vymezené, které mohou být použity pouze pro daný účel (např. .biz, .int, nebo .pro)
- Garantované, které mohou být použity pouze pro daný účel, navíc mají garanta – organizaci stanovující podmínky registrace a dohlížející na provoz. Označují se sTLD (podle sponsored TLD, kde sponsor znamená garant). Např. .aero .cat .coop .jobs .mobi .museum .travel
- Exkluzivní domény smějí používat pouze subjekty v USA, navíc s většími či menšími restrikcemi (např. .edu, .gov, .mil)
- Infrastrukturní domény jsou neveřejné (např. .arpa, .root)
- Atd.

Existují tyto generické TLD:

Struktura webové adresy



Katalog vs. vyhledávač

Pro vyhledávání v nekonečném množství informací na internetu lze využít dva typy vyhledávacích nástrojů:

- předmětové katalogy, neboli také Internet Directories, subject catalogs, subject directories, virtuální knihovny, či rozcestníky
- vyhledávací stroje, neboli také search engines či vyhledávače

Rozdíl mezi katalogy a vyhledávacími stroji je v tom, že data v katalogu jsou budována lidmi, kdežto data vyhledávačů jsou vystavěna na základě procházení internetových robotů sítí internetu.

Klasické katalogy (např. seznam.cz, centrum.cz, atlas.cz) lze přirovnat ke kartotékám v knihovně. Katalog je tvořen kategoriemi obsahujícími konkrétní záznamy vložené konkrétními osobami. Uživatel může procházet kategoriemi katalogu bez nutnosti vyhledávání podle klíčových slov. Není-li však záznam do katalogu nikým vložen, uživatel ho v katalogu nenalezne.

Katalog je pro vyhledávání informací na internetu vhodný hlavně v situaci, kdy uživatel neví přesně, jaký zdroj informací hledá, ale má přibližnou představu. Prohlíží si jednotlivé kategorie a pomocí hierarchických vztahů mezi kategoriemi a následně cestou od obecných položek ke specifitějším, může svůj požadavek specifikovat a dostat se k seznamu hodnotných zdrojů k požadovanému tématu.

Speciálním typem katalogů jsou tzv. virtuální knihovny (např. About.com, Britannica.com, Academicinfo.net, vlib.org), které se od běžných předmětových katalogů liší tím, že jsou vytvářeny odborníky a informace na nich jsou odborně tříděny a hodnoceny.

Většina katalogů je dnes obohacena o funkci vyhledávání pomocí klíčových slov.

Předmětové katalogy	
Výhody	Nevýhody
+ zpracované a popsané lidmi	- malý rozsah
+ kontrola kvality odkazů	- často nepřehledné řazení odkazů do kategorií
+ intuitivní navigace ve struktuře	- obsahují spíše obecná schémata
+ uvádí odkazy v dalších souvislostech	

Zdroj: (Boldiš, 2004)

Vyhledávací stroje (např. AltaVista, nebo Google) fungují na principu procházení odkazu po odkazu a organizaci získaných dat na základě textového obsahu procházených stránek. To je prováděno pomocí tzv. robotů, což jsou programy, které se pomocí hypertextových vazeb pohybují po webu a sbírají informace (indexují) o dokumentech a ukládají je do rozsáhlých databází, které pravidelně aktualizují. Vyhledávače procházejí zdroje na internetu a v případě shody s tzv. klíčovým slovem, které uživatel zadá do vyhledávače, sdělí, kde se požadovaný dokument nebo webová stránka nachází. Je třeba si však uvědomit, že zpracování dotazu a výsledky jsou automatické a nelze očekávat, že vyhledávač vždy porozumí hledanému smyslu slova. Největším problémem jsou proto vyhledané stránky, které sice obsahují hledané slovo, ale v jiném slova smyslu nebo i náhodný řetězec znaků.

Vyhledávací stroje	
Výhody	Nevýhody
+ široký záběr stránek (i několik miliard)	- hledá znaky ne význam slov
+ najde i specifické termíny	- kvalita odkazů kolísá
+ rychlost vyhledání	- množství odkazů
	- některé odkazy již nefungují

Zdroj: (Boldiš, 2004)

Základní pravidla pro vyhledávání

Zdroj: (JČU)

Základem úspěšného nalezení požadované informace je vědět, jak postupovat, kterou vyhledávací strategii a službu nebo nástroj zvolit. Dobrá strategie rovná se základ úspěšného vyhledání potřebných informací.

- formulujte slovy, co hledáte, jak moc má být informace aktuální, pro koho ji hledáte (profesionál, laik), jazykové omezení, geografické určení...
- správně zvolte klíčová slova a fráze (použijte odbornou terminologii), pozor na stop slova! Vhodnou kombinací klíčových slov se můžete pohodlněji a rychleji dobrat výsledku.
- používejte i synonyma, příbuzné výrazy, různé pravopisné tvary slov (vyhledávací nástroje pracují s informacemi v přirozeném jazyce)
- zkontrolujte pravopis klíčových slov
- zvolte vhodný typ vyhledávacího nástroje- lépe je vybrat několik vyhledávacích strojů (každý z nich může najít unikátní dokument)
- seznamte se s nápovědou (help) zvoleného vyhledávače
- sestavte dotaz pomocí klíčových slov a operátorů
- využijte možnost rozšířeného hledání
- vyhledaný výsledek vyhodnoťte
- zpřesněte formulaci nebo zkuste formulovat dotaz pomocí jiných klíčových slov a operátorů

Viditelný vs. neviditelný web

Internet lze rozdělit do dvou částí na tzv.:

- viditelný web, neboli také povrchový web, virble web či surface web
- neviditelný web, nebo také hluboký web, skrytý web, v angličtině invisible web, deep web, či hidden web.

Viditelný web je běžně volně přístupný a snadno indexovatelný běžnými vyhledávači.

Neviditelný web zahrnuje dokumenty, které jsou pro běžné vyhledávače obtížně vyhledatelné (neviditelné). Hluboký web představují typicky hodnotné stránky, které vyžadují autorizaci či předchozí registraci či prostě vše, co nelze prohledat prostými skoky z jednoho hypertextového odkazu na druhý. Jedná se především o informace uložené v databázích, adresáře, specializované vyhledávače, dokumenty v jiných formátech než HTML (např. PDF, PS, Macromedia Flash atd.), stránky tzv. samotáři, z kterých není odkaz na jiné a na které není odkaz, dynamicky generované stránky (po základní interakci s uživatelem)- např. katalogy knihoven, kalkulátory a také stránky chráněné heslem.

Hranice mezi viditelným a neviditelným webem je nejasná a postupem času se smazává, hlavně díky rychlému vývoji moderních vyhledávacích strojů, které přestávají být závislé pouze na obsahu stránek psaném v HTML kódu.

Velikost viditelného webu se odhaduje okolo 2-3 miliard indexovatelných dokumentů, zatímco velikost neviditelného webu se dá odhadnout přibližně na 400- 500 násobek webu viditelného.

Vzhledem k nemožnosti obsáhnout hluboký web běžnými vyhledávacími prostředky, je třeba využít k prohledávání webu specializované služby fungující na principu předmětových katalogů. Vyhledávání probíhá ve dvou krocích:

1. identifikace nejvhodnějších databází
2. položení konkrétního dotazu na zvolených databázích.

Specializované vyhledávací služby umožňující prohledávání neviditelného webu jsou např.:

- Complete Planet
- Invisible Web

- Direct Search
- Lycos Invisible Web Catalog
- Infomine Multiple Database Search
- AlphaSearch
- The Big Hub
- WebData
- Scirus

Google – operátory

Vyhledávání v Google si lze usnadnit pomocí tzv. operátorů. Následující tabulka obsahuje základní operátory a jejich vysvětlení.

„“	Vyhledá přesnou frázi
+	se umísťuje před slovo (bez mezery) a zaručí, že vaše výsledky budou obsahovat běžná slova, písmena či čísla, které vyhledávací technologie Google obvykle ignorují, jako např. v Karel IV.
-	se umísťuje před slovo (bez mezery) a vylučuje všechny výsledky, které obsahují tento vyhledávací termín, jako např. v [vrána -autor:vrána]
~	se umísťuje před slovo (bez mezery) a používá se v případech, kdy se mají vyhledat i synonyma k zadanému slovu
	zobrazuje výsledky, které obsahují alespoň jeden z vašich vyhledávacích výrazů
..	Používá se pro vyhledání číselného výsledku v intervalu dvou čísel [1900..1945]
*	Zástupný symbol pro jakékoliv slovo
OR	zobrazuje výsledky, které obsahují alespoň jeden z vašich vyhledávacích výrazů

Google podporuje několik pokročilých operátorů, což jsou specifická slova, které Google chápe zvláštním způsobem. Tyto operátory upraví zvláštním způsobem hledání, nebo se hledání provede jiným způsobem. Můžeme je chápat jako určité filtry, upřesňující výsledky vyhledávání.

Domain: určuje doménu, na které se bude hledat.

Site: Pokud při vyhledávání použijete operátor site: omezí Google vyhledávání na zadanou doménu.
Např: help site:www.google.com nalezne stránky o help v doméně www.google.com
help site:com nalezne stránky o help v doméně com

Filetype: Tento operátor nařizuje Google, aby vyhledával daný výraz jen uvnitř textových souborů specifického typu. Tento operátor potřebuje zadat ještě hledaný výraz jako argument.
Např.: "filetype:pdf sleva" vyhledá slovo "sleva" pouze uvnitř pdf souborů.
Podporované typy souborů, v kterých google dovede vyhledávat:

Adobe Portable Document Format (pdf); Adobe PostScript (ps); Lotus 1-2-3 (wk1, wk2, wk3, wk4, wk5, wki, wks, wku); Lotus WordPro (lwp); MacWrite (mw); Microsoft Excel (xls); Microsoft PowerPoint (ppt); Microsoft Word (doc); Microsoft Works (wks, wps, wdb); Microsoft Write (wri); Rich Text Format (rtf); SWF Shockwave Flash (swf); Text (ans, txt).

Link: vyhledá stránky, ze kterých vede odkaz na specifikovanou www stránku.
Např: link:www.google.com ukáže stránky, ze kterých vede odkaz na domovskou stránku Google.

Related: Pomocí operátoru related: vyhledáme stránky, které jsou "podobné" jako uvedená strana.
Např: related:www.google.com vyhledá stránky podobné úvodní stránce Googlu.

define: Dotaz define: vyhledá definice slov, která jsou uvedeny v různých zdrojích.
Např: define:web. K vyhledání definice se používá celá fráze (tzn. bude zahrnovat všechna použitá slova v přesně daném pořadí).

info: Dotaz info: ukáže informace, které má Google o specifikované web stránce.
Např: info:www.google.com zobrazí informace o domovské stránce Google.

- intext:** Umožňuje vyhledávání pouze v textu na samotné stránce, je ignorován titulek, veškeré hypertextové odkazy a internetové adresy. Google hledá v obyčejném textu. Použití tohoto příkazu je ideální právě pro výrazy, které se často v internetových adresách nebo v titulcích stránky objevují.
- intitle / allintitle:** Jestliže použijete operátor intitle / allintitle: Google omezí hledání na slova v titulu.
Např: allintitle:google search vyhledá pouze ty dokumenty, které mají v titulu google i search.
Např: intitle:google search vyhledá pouze ty dokumenty, které mají google v titulu, na umístění slova search nezáleží.
Pokud použijete intitle před každým slovem jedná se o alternativu k allintitle.
- inurl / allinurl:** Jestliže použijete operátor inurl: Google omezí hledání na slovo v URL.
např:inurl:google search vyhledá pouze ty dokumenty, které mají google v URL, na umístění slova search nezáleží.
Pokud použijete inurl před každým slovem jedná se o alternativu k allinurl.
Jestliže použijete operátor allinurl: omezí Google hledání na výskyt slov v URL.
Např: allinurl:google search vyhledá pouze ty dokumenty, které mají google i search v adrese stránky.
Všimněte si, že allinurl: pracuje se slovy a nikoli s částmi URL a že ignoruje interpunkci.
allinurl:foo/bar omezí hledání na výskyt slov "foo" a "bar" v URL, ale není nutné, aby tato slova byla v URL oddělena lomítkem, jak bylo zapsáno, stačí, pokud budou v zapsaném pořadí.
- cache:** Jestliže použijete operátor cache: Google nalezne toto klíčové slovo na stránce uložené v cache.
Např: cache:www.google.com web ukáže obsah uložené stránky www.google.com se zvýrazněným slovem web.
Dotaz cache: bez klíčového slova ukáže verzi web stránky, kterou má Google uloženou ve své paměti (cache).
Např: cache:www.google.com ukáže uloženou verzi úvodní stránky Google, která je ovšem shodná s původní "novou" stránkou.

Atd.

Text převzat a upraven z následujících zdrojů:

2009. *Wikipedie, otevřená encyklopedie.* [Online] 21. červen 2009. [Citace: 25. září 2009.] http://cs.wikipedia.org/wiki/Hypertext_Transfer_Protocol.

Boldiš, P. 2004. Pomůcky a manuály. *Pomůcky a manuály - Vzdělávání - SIC - Studijní a informační centrum.* [Online] 2004. [Citace: 15. září 2009.] <http://dl.webcore.czu.cz/file/ZGFRY0VseU83OU09>.

JČU. Jak vyhledávat informace na internetu. *Knihovny Jihočeské univerzity - Informační výchova.* [Online] [Citace: 15. září 2009.] http://www.lib.jcu.cz/vychova/vychova_jak5.htm.

Novák, M. 2006. SEO versus uživatelé a vyhledávače. *VOXCAFE - internet full service.* [Online] 3. prosinec 2006. [Citace: 15. září 2009.] <http://www.voxcafe.cz/clanky/optimalizace-stranek/seo-versus-uzivatele-a-vyhledavace.html>.

Velich, K. 2008. Vyhledávání na internetu a hluboký web. *Karel Velich.* [Online] 2. leden 2008. [Citace: 15. září 2009.] http://schacco.savana.cz/vlastni_web/zobrazit_prispevek.php?id=64.