

Název modulu:

Vyhledávací nástroje na internetu II

Cíl textu:

Doplnit animovanou přednášku o textovou formu a umožnit její tisk. Rozšířit základní znalosti o vyhledávacích nástrojích na internetu.

Základní pojmy:

- Vyhledávací stroj — vyhledávací nástroj, který umožňuje na základě zadaných klíčových slov vyhledat odpovídající informace na internetu
- Metavyhledávací stroj — vyhledávací nástroj, který přeposílá dotaz na jednotlivé vyhledávací stroje a vyhodnotí zároveň výsledky vyhledávání od jednotlivých vyhledávacích strojů na jednom místě
- Neviditelný web (invisible web) — hluboký web (deep web)

1 ÚVOD MODULU

Vyhledávací nástroje na internetu jsou výrazným pomocníkem při orientaci i při vyhledávání v síti sítí. V modulu Vyhledávací nástroje na internetu I jsme si představili různé druhy vyhledávacích nástrojů, nyní se podíváme na některé z nich prakticky. Konkrétně si představíme inteligentní nástroje na vyhledávání, jakými jsou především vyhledávací stroje (search engines) a metavyhledávací stroje (metasearch engines). Pozastavíme se u vyhledávacích nástrojů, které proniknou do hluboko ukrytých částí webu. S vyhledáváním na internetu souvisí i téma evaluace vyhledaných informací, kterou bychom neměli podceňovat.

2 KDE SE DOVÍM, JAKÉ VYHLEDÁVACÍ NÁSTROJE EXISTUJÍ?

Pro efektivní vyhledávání na internetu můžeme využívat desítky vyhledávacích nástrojů. Velmi praktické jsou souhrnné rozcestníky, adresáře vyhledávacích nástrojů, kde se můžeme zorientovat, který z nástrojů použít. Příkladem takového adresáře může být:

- Beaucoup - <http://www.beaucoup.com/>
- SearchengineCollossus - <http://www.searchenginecollossus.com/>

Kde číst o
vyhledávacích

Řada specializovaných zdrojů sumarizuje ve srovnávacích studiích o vyhledávacích strojích. Nalezneme zde i aktuální informace o nových vyhledávacích funkcích – např:

- SearchengineWatch - <http://searchenginewatch.com>
- SearchengineShowDown - <http://searchengineshowdown.com/>
- Top SearchEngines <http://topsearchengines.info/>

2.1 VYHLEDÁVACÍ STROJE: GOOGLE A TI DRUZÍ

Proč se namáhat s učením a využíváním jiných vyhledávacích nástrojů, když máme Google? Tento vyhledávač velmi změnil situaci na poli vyhledávacích strojů a jistě ho není potřeba podrobně představovat. V čem spočívá jeho veliká popularita na celém světě a náskok před konkurenčními stroji? Můžeme shrnout alespoň některé kladné body pro Google:

- originální nápady, služby, nástroje
- rozsah prohledávaného webu (resp. velmi rozsáhlá databáze)
- relativní přesnost vyhledávání
- uživatelská přívětivost

2.1.1 Zajímavé funkce, služby Googlu

Známe opravdu všechny jeho kvalitní vyhledávací nástroje a funkce? Nejprve si připomeneme užitečné rady a techniky pokročilého vyhledávání.

Praktický příklad

2 a více pojmů defaultně spojuje operátorem AND (vyhledá stránky, kde figurují všechna zadaná hesla)

Star Wars Episode +I - „I“ musí být ve vyhledaných stránkách obsaženo (patří mezi takzvaná „stop slova“)

“léčivé rostliny” - najde přesnou frázi

cembalo OR virginal - najde stránky s minimálně jedním termínem

Klaus -Santa označení „Santa“ nesmí být ve vyhledaných stránkách obsaženo

„* ze Žerotína“ – hvězdička nahradí libovolné slovo

~copyright - hledá uvedený termín včetně synonymních výrazů, např. i intellectual property. (Zatím lze použít pouze pro slovní zásobu v angličtině).

Pokud by vám připadaly googlovské pokročilé techniky překombinované, můžete vyzkoušet stejné funkce v obrázkovém a přívětivém prostředí na: <http://www.soople.com>

Pro efektivní vyhledávání můžeme použít i speciální typ vyhledávání formou určitých operátorů (příkazů). Způsob zápisu dotazu ve vyhledávacím okně je vždy ve formě:

příkaz:zadání k vyhledávání (mezera) další termín

Praktický příklad:

filetype:pdf "digitální knihovny,, - omezení jen na určitý formát dokumentu (PDF, XLS, DOC, PHP...)

intitle:CPU Athlon - najde „CPU“ v názvu stránky a „Athlon“ kdekoli

allintitle:idnes galerie - najde „idnes“ a „galerie“ v názvu stránky (současně)

link:www.lupa.cz - najde stránky obsahující odkaz na stránky uvedeného zdroje

allinanchor:Oslo - najde stránky obsahující „Oslo“ v odkazu

related:http://www.zaskolou.cz - najde tématicky podobné stránky k uvedené adrese

site:www.micr.cz „elektronický podpis“ - na uvedené www prohledá zdroje na téma elektronický podpis

inurl:knihovna – najde uvedené slovo v URL adrese

define:manuscript – najde definici pojmu

stocks:IBM – najde burzovní, firemní informace o firmě

info:www.muni.cz – vypíše informace o stránce (cache, odkazy na uvedenou stránku)

Zajímavým doplňkem pro vyhledávání je tzv. **Google Desktop**. Je to volně stažitelný program, který slouží k prohledávání lokálních souborů a adresářů v různých programech (ve vašem počítači). Desktopy nabízí i řada dalších firem, vyhledávacích strojů (např. Yahoo).

Dalšími zajímavostmi z dílny Google jsou:

Google
vychytávky

Google Scholar <http://scholar.google.com/>

- prohledávání odborných informací – vědeckých článků, preprintů, technických zpráv, vysokoškolských kvalifikačních prací
- pokud není volně dostupný plný text, zobrazí se alespoň abstrakty příspěvků
- pokročilé vyhledávání je možné omezit oborově

Google Book Search <http://books.google.com/>

- zjišťování základních informací o vyšlých knižních titulech (recenze, vydavatelské údaje, odkazy na online obchody)
- možnost prohlížení údajů o knize v knihovních katalozích
- prohlížení úryvků z textů knih (pokud není konflikt s autorskými právy)

2.1.2 Čím mohou konkurovat Googlu jiné vyhledávací stroje?

Google není jediný vyhledávací stroj. Mezi další srovnatelné vyhledávací stroje patří např.:

Ask - <http://www.ask.com> (zajímavé je např. použití přirozeného jazyka - mluvené řeči)

Alltheweb - <http://www.alltheweb.com> (zajímavé je např. dokončování slov, nabízená struktura a souvisejících témat)

Altavista - <http://www.altavista.com/> (zajímavá je např. možnost překladů textů, celých www stránek Babel fish translation)

Exalead - <http://www.exalead.com/search> (zajímavé je např. řešení zobrazení výsledků)

Specializované vyhledávací nástroje encyklopedického a slovníkového typu:

Answers.com - <http://www.answers.com/>

One Look: dictionary search - <http://www.onelook.com>

Vyhledávání s „lidským rozměrem“

Řada internetových služeb se snaží dnes lidem vytvořit prostředí, kde jim umožní spoluvytvářet a sdílet informace novým sociálním způsobem např. díky využití wiki, blogů a dalších komunikačních nástrojů.

Konkrétní ukázky použití wiki (spoluvytváření obsahů) můžete vidět na nejznámějším projektu – otevřené encyklopedii Wikipedie, česky na: <http://cs.wikipedia.org>.

Pro vyhledávání blogů (deníčky, zprávy) nejruznějšího zaměření můžete využít specializované zdroje jako např.:

Technorati <http://www.technorati.com>

Feedster <http://www.feedster.com/>

Blogportál <http://blogportal.hlava.net/>

V současné době nechtějí být uživatelé internetu jen pasivními strávníky informací na internetu, ale chtějí se aktivně podílet na tvorbě obsahů, hodnocení a popisu stránek apod. Vyhledávací nástroje, které jsou postaveny na „společném sdílení a hodnocení odkazovaných stránek“, jsou např.:

Del.icio.us <http://del.icio.us/>

CiteULike <http://www.citeulike.org>

Trendy

2.2 METAVYHLEDÁVACÍ STROJE

Metavyhledávače principiálně pracují odlišně než vyhledávací stroje. Po zadání dotazu prohledávají velké množství Internetu - dotaz je předáván k vyhodnocení konkrétním rešeršním systémům (Googlu, Altavistě...).

Clusty <http://clusty.com/> (navigace podle kategorií=klastrů v bočním menu)

Dogpile <http://www.dogpile.com/> (přehledné zobrazení výsledků vyhledávání z konkrétních vyhledávacích strojů, navigace podle témat v levém sloupci)

Kartoo <http://www.kartoo.com/> (efektní vizualizace)

3 VYHLEDÁVACÍ NÁSTROJE HLUBOKÉHO WEBU

Zatím je z principu nereálné, aby existoval vyhledávací nástroj, který by uměl prohledávat úplně všechny informace na webu.

Hluboký, neviditelný web je označení pro takové stránky a zdroje na internetu, které nenajdeme běžnými vyhledávacími nástroji (např. Google). Neviditelný web tvoří z největší části informace skryté ve specializovaných databázích.

Určitě stojí za to prozkoumat i tato bílá místa. Neviditelný web je údajně až 500krát obsáhlejší než web povrchový = prohledatelný běžnými vyhledávacími nástroji.

Mezi nástroje, kterými se můžeme podívat pod pokličku neviditelného webu, patří např.:

Complete Planet <http://aip.completeplanet.com>

- registruje přes 70 000 rozsáhlých specializovaných zdrojů (databází, vyhledávacích strojů neviditelného webu)

Turbo 10 <http://turbo10.com/>

- metavyhledávací nástroj neviditelného web
- zajímavá vizualizace, třídění=clustrování výsledků vyhledávání podle kategorií
- obsahuje vládní, univerzitní, obchodní online zdroje, databáze

Konkrétní systémy na prohledávání hlubokého webu
--

Scirus <http://www.scirus.com/>

- přes 250 mil. vědecky zaměřených www stránek
- specializovaný vyhledávací nástroj pro vědecké informace (články, preprinty, patenty, výzkumné zprávy)
- náměty na vyhledávání <http://www.scirus.com/srsapp/tips/>

Looksmart: find article <http://www.findarticles.com/>

- přístup k volně dostupným článkům i placeným zdrojům od prestižních vydavatelů

Portál vědeckých informací **Infoplease** <http://www.infoplease.com/index.html>

Akademický portál **Infomine**: <http://infomine.ucr.edu/>

- obsahuje primární dokumenty z různých oblastí vědy, vzácné tisky, prvotisky, historické fotografie

4 HODNOCENÍ INFORMACÍ

Při vyhledávání na internetu velice často narazíme na problém, jak oddělit zrno od pleve. Internet je úžasně svobodný prostor, kde se může vyjádřit kdokoli o čemkoli. Na internetu nemáme žádného cenzora obsahů. Řada informací žije vlastním životem a nikdo je neaktualizuje. Musíme tedy být připraveni kritickým pohledem posuzovat výsledky, které nám zobrazí libovolný vyhledávací nástroj. S hodnocením obsahu nám nepomohou ani sofistikované vyhledávací stroje, protože nijak nehodnotí obsahovou kvalitu nalezených www stránek.

Každý očekáváme od výsledků vyhledávání něco jiného, a co se nelíbí jednomu, druhý považuje za přínosné. Evaluace informací bude do jisté míry vždy subjektivní, ale některá obecná kritéria pro nás mohou být užitečným vodítkem při hodnocení výsledků vyhledávání na internetu. Určitě nám mohou ušetřit sílu a čas.

Jak hodnotit výsledky vyhledávání

Kritéria hodnocení:

- **kritérium - jasný účel textu (práce-vědecký text či reklama), cílový uživatel, uvedení důvodů pro zveřejnění**
- **kritérium úplnosti, přesnosti, důvěryhodnosti, objektivita, podložená argumentace u odborných textů**
- **kritérium - jasné autorství**
- **jazyk textu, stylistika**
- **umístění adresy, doména může napovědět více o autorovi, příp. instituci**
- **u článků, odborných statí by měla být uvedena citovaná literatura, odkazy**
- **datum zveřejnění, aktualizace**
- **možnost ověření v referenčních zdrojích**

Pro posuzování kvality obsahu konkrétní www stránky mějte na paměti sadu kontrolních otázek, která vám usnadní jejich hodnocení

Desatero kritických otázek než klikneme na vyhledanou www stránku

Kontrolní desatero:

- 1. Kdo je autorem obsahu?**
- 2. Jakou má autor kvalifikaci, odbornou erudovanost ve vztahu k popisovanému tématu?**
- 3. Jsou prezentované informace přesné, objektivní, přiměřené?**
- 4. Je www stránka po obsahové stránce srovnatelná s jinými stránkami zabývajícími se podobným tématem?**
- 5. Obsahuje www stránka nové, originální myšlenky, informace k tématu a jsou tyto informace podloženy objektivními argumenty?**
- 6. Odpovídá styl jazyka zamýšlenému cílovému publiku?**
- 7. Můžeme zjistit reference, odkazy na www stránku nebo autora informací?**
- 8. Můžeme zjistit datum publikování a poslední aktualizace www stránky?**
- 9. Obsahuje www stránka funkční linky, odkazy?**
- 10. Je www stránka pro navigaci přehledná a případné obrázky, videa, zvuky mají opodstatněnou funkci na stránce?**

4.1 MOŽNOST SLEDOVÁNÍ KLAMAVÝCH STRÁNEK A PODVODŮ

Některé informace se tváří seriózně, ale jsou to jen zlomyslné žerty nebo mystifikace. Lépe poznat některé z poplašných zpráv, které kolují na internetu serveru, můžeme např. na serveru **HOAX** - <http://www.hoax.cz/>

Z mezinárodního prostředí si můžete prohlédnout stránky institucí:

IFCC – Internet Crime Complaint Centrum – <http://www.ic3.gov>

- partnerská instituce FBI, shromažďuje informace na podvodné služby a zdroje na internetu, registr stížností + zveřejnění aktuálních hrozeb

FRAUD.org – <http://www.fraud.org/internet/inttip/inttip.htm>

- národní informační centrum, které monitoruje podvody na internetu.