
Rešeršní činnost

Vyhledávací nástroje internetu 2

25. 4. 2008, 2. 5. 2008

přednášející: Silvie Kořínková Presová
presova@phil.muni.cz
Kabinet inf. studií a knihovnictví, FF MU

Vyhledávací stroje

V čem se liší od předmětových adresářů?

☞ mnohem větší databáze

☞ automatizovaný sběr dat

☞ zaměřeny na vyhledávání spíše než na prohlížení

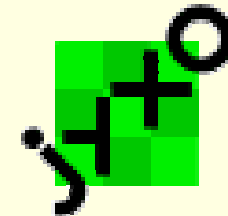
Vyhledávací stroje

všeobecně zaměřené vyhl. stroje

- Yahoo!
- Google
- Ask
- Live Search
- Exalead

české

- Morfeo
- jyxo
- Seznam

The logo for Yahoo!, featuring the word "YAHOO!" in a bold, red, sans-serif font with a slight shadow effect.The logo for Google, featuring the word "Google" in its characteristic multi-colored font (blue, red, yellow, blue, green, red) on a white background.The logo for Ask.com, featuring the word "Ask" in a white, bold, sans-serif font inside a red oval, with ".com" in a smaller font below it and a trademark symbol (TM) to the right.

Vyhledávací stroje

Je výhodné použít, když:

- ☞ hledáme specifické téma, předmět hledání je úzký
- ☞ chceme prohledávat milióny webových stránek
- ☞ chceme vyhledat velký počet záznamů k předmětu zkoumání
- ☞ hledáme určité typy dokumentů/informací, souborů
- ☞ chceme použít příkazy pro rychlé vyhledávání, omezení pomocí filtrů, jiné možnosti pokročilého vyhledávání

Vyhledávací stroje

dvojí typ rozhraní pro vyhledávání

- jednoduché rozhraní

zadávaní jednoduchých dotazů, pokročilé dotazy – nutné znát dotazovací jazyk

- rozhraní pro pokročilé vyhledávání –
návodné menu

Výsledky vyhledávání jsou tak dobré, jak kvalitní je dotaz uživatele. (Sklenák, 2001)

Nástroje pro vyhledávání – formulace dotazů

AND (+) , OR, NOT (AND NOT, -) (bool. operátory)

- ❑ **AND** – současný výskyt obou vyhledávacích termínů, zužuje dotaz
- ❑ **OR** – alespoň jeden z vyhledávacích termínů, rozšiřuje dotaz
- ❑ **NOT** – operátor vyloučení dokumentů obsahující zadaný termín

proximitní operátory: operátor, který specifikuje pořadí anebo vzdálenost mezi dvěma vyhledávacími slovy, např. **NEAR** (v **Exalead** 16 slov, ale dá se měnit např. women NEAR/3 politics, **Morfeo**, **Google** tomáš * masaryk), více viz [Distanční operátory](#)

fráze “ ” - řetězec slov, který se v textu dokumentu musí vyskytovat v přesně stanoveném pořadí vedle sebe

- ☞ slouží pro specifikaci, vyhledávání ustálených sousloví a spojení
- ☞ pokud nezáleží na pořadí použijte vhodné prox. operátory či AND

truncation – krácení podle slovních kořenů *

stemming – výklad viz TDKIV, např. jyxó, Google

wildcards – zástupné znaky ? *

fullword wildcard – nahrazuje celé slovo, např. Google *

- ☞ slouží pro formulování širšího dotazu, rozšíření dotazu
- ☞ vyjádření jednotného i množného čísla (**knihovn***), pravopisné možnosti (**disku?e** – diskuse OR diskuze)
- ☞ mohou být vyhledány i významově odlišná slova – **hlad***, **hlad'**, **hladina**, **hladký**

Pravidlo tří konceptů

☞ doporučuje se „položit“ vedle sebe tři pojmy

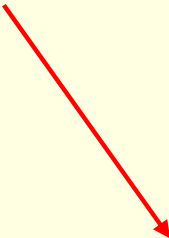
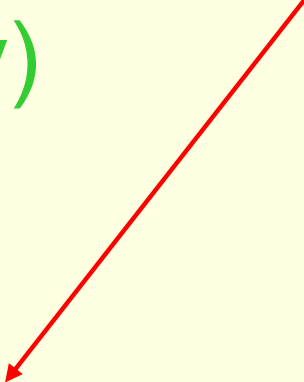
„rešeršní služby“ 

první koncept

reprezentuje hlavní předmět zájmu
(odpovídající rozlišovací úroveň)

AND (knihovna OR „informační instituce“)

AND (ceník OR ceníky)

  další dva koncepty by měly odpovídat
doplňkovým hlediskům např. o místu,
příčině apod.

Použití kulatých závorek

změna pořadí pro vyhodnocování – to, co je v závorce, se vyhodnocuje jako první

princip „zevnitř-ven“

- ☞ pojmy, které mají být vyhodnoceny jako první, musí být na nejnižší úrovni vnoření

Třetí výraz (druhý výraz (první výraz k vyhodnocení) k vyhodnocení)

- ☞ výrazy na stejné úrovni – vyhodnocovány zleva doprava, přednost má operátor AND před OR

Příklad

- Ask nepodporuje závorky
- Yahoo!, Google - nejsou nutné
- Live Search - operátor OR

(rešeršní služby OR rešerše) knihovna

(rešerše OR rešeršní služby) knihovna

zkus na www.live.com

For this operation

Try this on Google

$x \text{ AND } y$

$x \ y$

$x \text{ OR } y$

$x \text{ OR } y$

$x \text{ AND } (y \text{ OR } z)$

$x \ y \text{ OR } z$

$(x \text{ AND } y) \text{ OR } (z \text{ AND } q)$

not yet possible

$(x \text{ OR } y) \text{ AND } (z \text{ OR } q)$

$x \text{ OR } y \ z \ \text{OR } q$

$x \text{ AND } (y \text{ OR } z \ \text{OR } q)$

$x \ y \ \text{OR } z \ \text{OR } q$

$x \text{ AND } (y \text{ OR } z) \ \text{AND } q$

$x \ y \ \text{OR } z \ q$

Použití filtrů

vyhledávání podle pole – v celém url, omezení na doménu, název stránky

formální hlediska pro zpřesnění dotazu – jazyk, typ dokumentu

omezení prostoru dokumentů, v němž je dotaz vyhodnocován

- zadání do formuláře (menu v pokročilém rozhraní)
- ze základního rozhraní pomocí definovaných příkazů
- zpřesnění dotazu

Filtr uzlu

- ☞ umožňuje omezit prostor vyhledávání na jeden konkrétní uzel nebo specifickou podmnožinu uzlů podle struktury URL (vyjadřuje adresu umístění libovolného IZ na internetu)
- ☞ porozumění částem URL ⇒ ⇒ efektivní používání filtru uzlu (SKLENÁK, 2001:292):

http://www.inforum.cz/inforum2004/pdf/Peceny_Ondrej.pdf

standardní
prefix

jméno
subdomény

doména nejvyšší
úrovně

umístění daného
dokumentu v
rámci uzlu

Příklady příkazů v



- ➡ **inurl:radioservis eco** - nalezne slovo „radioservis“ v URL a „eco“ kdekoli na stránce
- ➡ **allinurl:phil rozvrh** - nalezne slovo „phil“ a zároveň „rozvrh“ v URL
- ➡ **site:rozhlas.cz rok v dopisech** - omezení vyhledávání dokumentů pouze na uvedenou doménu

Příklady příkazů v



- **intitle:"iva bittová" viklický** - nalezne uvedenou frázi v názvu stránky a „viklický“ kdekoliv na stránce
- **allintitle:rešeršní strategie** - nalezne termín „rešeršní“ a zároveň „strategie“ v názvu stránky
- **inanchor:dialog databáze** - nalezne slovo „dialog“ v odkazu a „databáze“ kdekoli na stránce
- **allinanchor:digitální knihovna** - nalezne slovo „digitální“ a zároveň „knihovna“ v odkazu
- **link:www.stk.cz** - nalezne stránky s odkazem na stránky Státní technické knihovny
- **related:www.ipl.org** - nalezne stránky podobné uvedené stránce

☞ Každý vyhledávací stroj má své specifické rysy – odlišné dotazovací jazyky

☞ Search Engine Features Chart

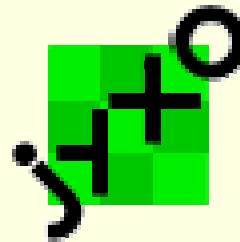
☞ Search Engines by Search Features

☞ Weare, W. H., 2008

☞ nápovědy u jednotlivých vyhledávačů:
Advanced Search Tips, Web Search Help,
Help and How to Search aj.

Možnosti zúžení dotazu

Některé vyhledávací stroje nabízejí uživateli pokročilé možnosti, jak výsledek vyhledávání zúžit



Různé možnosti práce s dotazem

- ☞ Yahoo! - [Search Assist](#) – navrhuje další zpřesnění dotazu
- ☞ Ask - výsledný dotaz – nabídka vlevo: [Narrow Your Search](#), [Expand Your Search](#), vpravo – výsledek dle typu informace (encyklopedie, obrázky, rss)
- ☞ Google - na konci stránky - [Searches related to](#), [hledání ve výsledcích](#)
- ☞ Exalead - různé možnosti [Narrow your search](#)

Vyhledávání nejrozmanitějších obsahů

- 👉 image
- 👉 news
- 👉 desktop
- 👉 weblog
- 👉 maps
- 👉 videos
- 👉 encyclopedia
- 👉 shopping and product search
- 👉 rss
- 👉 audio, podcast

Groups

Diskusní skupiny (angl. discussion groups, newsgroups) - umožňují lidem komunikovat na různá témata, dva způsoby komunikace:

- ☞ výměna informací na speciálních stránkách – online forum
- ☞ výměna informací v rámci e-mailové skupiny – mailing list, elektronické konference

Google Groups

Vyhledávání Usenet groups a Google Groups, které nejsou součástí Usenet

- ➡ procházení tématy
- ➡ přímé vyhledávání diskusních skupin
- ➡ možnost pokročilého vyhledávání

- ➡ existuje počestěná verze

Služby pro vyhledávání groups

- ☞ Yahoo Groups
- ☞ Delphi Forums - více než 100 tis.aktivních skupin
- ☞ Big Boards – adresář zaměřený na kategorizace disk. skupin na webu

Mailing list

- ☞ Topica – služba pro správu elektronických konferencí, možnost vyhledávání, kategorizace
- ☞ L-Soft CataList, the Official Catalog of LISTSERV® lists – oficiální katalog konferencí používající Listserv software

Google Books

- ☞ vyhledávání nových i starších knih
- ☞ úmluva s vydavateli knih a knihovnami
- ☞ bibliografické údaje x plné texty
- ☞ existuje počestěná verze - <http://books.google.cz/>

Různé typy zobrazení:

- ☞ úplné zobrazení - Full view
- ☞ omezený náhled - Limited preview
- ☞ zobrazení fragmentů - Snippet view
- ☞ náhled není k dispozici - No preview available

Plné texty knih online

- [The Online Books Page](#)
- [Project Gutenberg](#)
- zpřístupňují knihy, které nepodléhají autorskému právu

Historické dokumenty

- [EuroDocs:Primary Historical Documents From Western Europe](#)
- [A Chronology of US Historical Documents](#)
- [University of Virginia Hypertext Collection](#)

Vyhledávání obrázků

- ☞ databáze obrázků vyhledaných na webu vyhledávacími stroji – Google (1,8 biliónů obrázků), Yahoo! (1,6 biliónů obrázků), Ask
- ☞ specializované kolekce obrázků – tematické konkr. organizací, zprostředkujících organizací

Vyhledávání obrázků

☞ neustálý vývoj technologií

Co indexují vyhledávací stroje?

☞ jméno souboru (např. krajina.jpg)

☞ tag Alt – text asociovaný s obrázkem, objevuje se při najetí myši na obrázek

☞ text poblíž obrázku – zapříčiňuje nerelev. výsledky

☞ vyhledávání obrázků – volba max. dvou vyhledávacích výrazů

Databáze obrázků

Adresáře/seznamy databází obrázků

- [Digital Librarian: a librarian's choice of the best of the Web: Images](#)
- OIB - [BUBLLINK: Image Collections](#)

Komerční sbírky kvalitních obrázků

- [Corbis](#) – různé druhy grafiky – fotografie, obrázky, ilustrace
- [FotoSearch](#)
- [Creative Commons](#) – vedle fotografií registruje audio, video, text a výukové materiály

Ukládání a sdílení fotografií

- [Flickr](#)

Zvukové nahrávky

Adresáře

- ➔ World Wide Web Virtual Library: Audio –
kategorie general repositories, newsgroups,
online radio, software
- ➔ Digital Librarian: A Librarian's Choice of the
Best of the Web - Audio

Zvukové nahrávky

Vyhledávací stroje

Yahoo! Audio Search

vyhledávání hudby, zpravodajství, rozhovorů, jiné mluvené slovo, podcastů aj.

Singingfish

vyhledávání audio, video

Podcast

- ☞ technologie podcast je založená na RSS exportech, které obsahují popis audionahrávky a odkaz na místo, kde je soubor uložen
- ☞ speciálně zformátovaný soubor nabízí pravidelně aktualizovanou nabídku audiosouborů, které jsou na příslušné stránce k dispozici
- ☞ speciální program

Podcast České rozhlasu

Podcast

- ☞ Yahoo!'s Podcast Search – součástí Yahoo!
Audio
- ☞ Podcastdirectory.com – předmětový adresář
- ☞ Podscope – vyhledávání audio i video,
využívá vlastní technologii na rozeznávání
mluveného slova a využívá ji pro vyhledávání

Video

- ☞ BUBLLINK / 5:15 Catalogue of Internet Resources: Video – odkazy na webová sídla, která zpřístupňují videa či odkazují na další zdroje
- ☞ Yahoo! Video Search – milióny videí sesbíraných roboty, vedle toho videa získaná přímo od vydavatelů/tvůrců videí prostřednictvím RSS
- ☞ Google Video Search
- videa uložena na serverech G. (videos hosted by Google), poskytnutá ke sdílení, YouTube
- ☞ YouTube – vystavování a hledání videí

Vyhledávání blogů

- Bloglines – vlastní Ask (lze vyhledávat i přes Ask), vyhledávání RSS
- Technorati – vedle vyhledávání i kategorizace
- Google Blog Search – mapuje i český web, pokročilé vyhledávání

RSS

- Bloglines
- Live Search – výběr Další – Informační kanály, zmapován i český internet
- Google - přestal zahrnovat do výsledku vyhledávání RSS, protože jeho algoritmus řadil tyto typy informací nejvýše

Co vše lze vyhledávat?

- ☞ [Top 25 Web 2.0 Search Engines](#)
- ☞ [Top 100 Alternative Search Engines](#)
- ☞ [Intelways](#)
- ☞ [The Extreme Searcher's Internet Handbook](#)

Metavyhledávací stroje

Umožňují uživateli na základě jednoho dotazu paralelní prohledávání databází několika vyhledávacích služeb (TDKIV)



zadávání dotazu pouze jednou, ten je pak paralelně vyhodnocen

Výhody

- kombinování výsledků vyhledávání s odstraněním duplicity
- jednotné uživatelské prostředí
- uživatelé nemusí sledovat nově se objevující vyhledávací systémy
- vyšší úplnost vyhledávání, možnost získat více relevantních dokumentů

Metavyhledávací stroje

Nevýhody

- **limitují počet záznamů z vyhledávacího stroje**
 - celkový výsledek sestavuje z 10 – 50 hitů převzatých z každého vyhledávače
- **ztráta pokročilého vyhledávání, sofistikovanějšího formulování rešeršního dotazu**
- **většinou využívají 2 – 3 z nejvýznamnějších vyhledávačů**
- **nízká výkonnost, odezva**

Metavyhledávací stroje

- <http://www.dogpile.com/>
- <http://clusty.com/> - shlukování dokumentů
- <http://www.kartoo.com/>

typu **all-in-one** – webové sídlo se seznamem vyhledávacích nástrojů, ne paralelní vyhledávání

- <http://www.intelways.com/>
- <http://www.globalsearch.cz>

Vyhledávací strategie

... se zaměřením na vyhledávací stroje

(SKLENÁK, 2001:306)

- ✓ obecné vyhledávání
- ✓ specifické vyhledávání
- ✓ inkrementální vyhledávání
- ✓ řetězcové vyhledávání
- ✓ vyhledej a skoč
- ✓ kategoriální vyhledávání

Obecné vyhledávání

uživatel má obecnou představu o tom, co chce najít

☞ nezná podrobnosti

vyhledání do šíře

prostředkem pro obecné vyhledávání je operátor

OR

☞ umožňuje vyhledávat dokumenty, které obsahují libovolné z klíčových slov

☞ vhodné použít při různých tvarech nebo synonymech

nebezpečí v příliš obecných dotazech

☞ zajímavé informace jsou „pohřbeny“ v návalu nerelevantních dokumentů



uživatel hledá něco o službách
knihoven

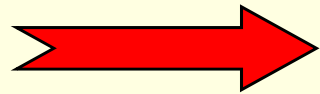
knihovní služby OR služby knihoven OR
knihovnické služby OR informační služby

Specifické vyhledávání

Pokud uživatel zná podrobnosti k tématu



je vhodné použít tento způsob



je založeno na operátoru AND,

pomocí kterého jsou kombinovány různé pojmy

nejen pomocí AND $\Rightarrow \Rightarrow$ využití filtrů,
proximitních operátorů apod.

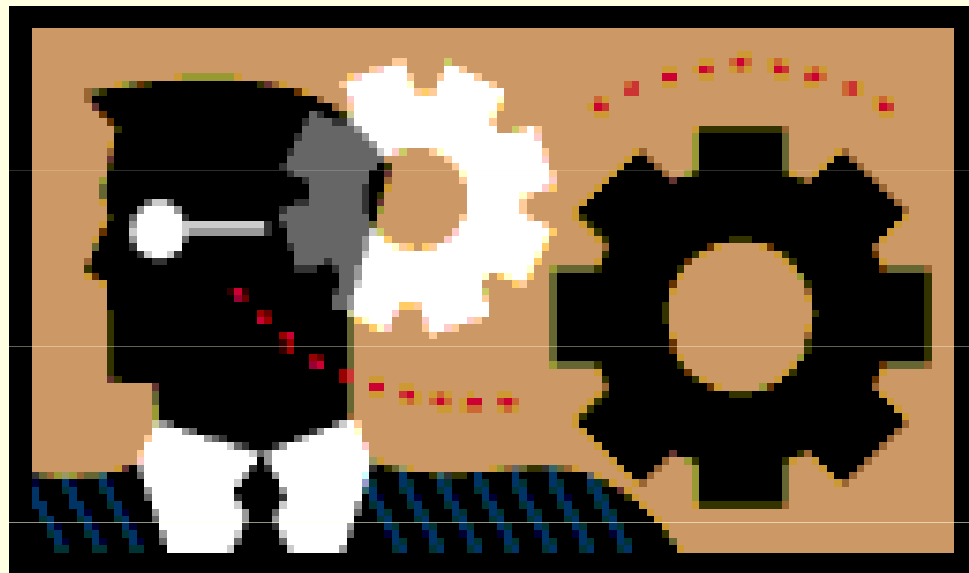


doporučuje se začít tímto vyhledáváním,
pokud uživatel dokáže v jisté míře
specifikovat svůj požadavek, po neúspěchu

\Rightarrow zobecnění dotazu

☞ při kombinování dotazů je třeba dávat si pozor na kombinování a pořadí klíčových slov a dotazů

PROČ?



požadavek: *Rešeršní služby knihoven*

*rešeršní služby rešerše knihovna informační
knihovny instituce* .



Rešeršní služby knihoven

- knihovna OR („informační instituce“ AND „rešeršní služby“)

ve výsledcích je požadováno **alespoň slovo knihovna** anebo **současný výskyt slov informační instituce a rešeršní služby** současně anebo **všechny tři současně**

- „rešeršní služby“ AND (knihovna OR „informační instituce“)

ve výsledcích **se musí vyskytovat spojení rešeršní služby a alespoň jedno ze slov knihovna či informační instituce**

Inkrementální vyhledávání

Pokud nejsme spokojeni s výsledkem vyhledávání ⇒ ⇒ změnit dotaz a zkusit to znovu

- ☞ obvyklým postupem při tomto vyhledávání je postupné zpřesňování dotazu, tzv. *ladění dotazu*, tj. maximální přiblížení formulace skutečné informační potřebě uživatele
- ☞ využívání operátorů **AND**, **OR**, **NOT**
- ☞ zpřesňování pomocí formálních hledisek, filtrů, tj. udání polí, kde se má vyhledávat

„rešeršní služby“ AND (knihovna OR
„informační instituce“)

⇒ omezení na lékařské knihovny

⇒ ⇒ využij v



funkce „Search within results“

Řetězcové vyhledávání

- ❑ většina vyhledávacích strojů hledá pouze celá slova
 - ⇒⇒ knihovna knihovny knihoven
 - ⇒⇒ použití OR

- ❑ některé vyhledávací stroje mohou vyhledávat i části slov neboli řetězce (nebo přesněji podřetězce) ⇒⇒ stemming
 - ❑ Librarians' Internet Index library libraries

- ❑ riziko nerelevantních hitů
 - ❑ library science v LCSH
 - ❑ Military art and science
 - Military libraries → libraries science

Vyhledej a skoč

- ☞ vyhledávání v hitech na zobrazené stránce pomocí FIND
- ☞ spočívá v tom, že uživatel se pomocí klíčových slov rychle orientuje v seznamu vyhledaných odkazů na dokumenty
 - to předpokládá zobrazení maxim. počtu hitů na stránku
- ☞ nevýhody:
 - ❑ delší odezva při zadání základního dotazu
 - ❑ jednoduchost vyhledávání – ne složené podmínky (vyhledávání jen podle řetězce)
 - ❑ prohledává pouze text, který prezentuje hity

Kategoriální vyhledávání

- ☞ navigace hierarchickou strukturou
předmětových adresářů/katalogových
vyhledávacích nástrojů
 - ☐ prohlížení jednotlivých úrovní v příslušné
kategorii
 - ☐ oborové brány **versus** komerční adresáře
- ☞ posun od obecnějších kategorií
k specifitějším (záleží na tom, jaké nabízí
služba propracované schéma)

Povinná literatura

- ☞ Antoš, D. **Google přichází s průlomovými změnami ve vyhledávání.** *Lupa* [online]. 21. 5. 2007 [cit. 2008-04-27]. Dostupné na World Wide Web: <http://www.lupa.cz/clanky/google-prichazi-s-prulomovymi-zmenami-ve-vyhledavani/>
- ☞ Hock, R. **Search Engines : From Web 0.0 to Web 2.0 and Beyond.** *Online*, 2007, roč. 31, č. 1, s. 26-30.
- ☞ Weare, W. H. **Find It on the Web Using the Search Concepts You Already Know.** *Library Media Connection*, 2008, roč. 26, č. 6, s. 56-58.
- ☞ Papíková, V., Papík, R. **Nové možnosti vyhledávání vědeckých lékařských informací.** In *INFORUM 2007 : 13. konference o profesionálních informačních zdrojích Praha, 22. - 24. 5. 2007 : sborník z konference informačních profesionálů* [online]. Dostupné na World Wide Web: <http://www.inforum.cz/pdf/2007/papikova-vendula.pdf>

Doplňující a použitá literatura

- ☞ Hock, R. **The extreme searcher's Internet handbook : a guide for the serious searcher**. 2nd ed. Medford, N.J. : CyberAge Books, 2007. 326 s.
- ☞ Iskra, J. **Google : tipy a návody pro vyhledávač, Gmail, YouTube, Earth a další aplikace**. Brno : Computer Press, 2008. 231 s.
- ☞ Makulová, S. **Vyhľadávanie informácií v internete : problémy, východiská, postupy**. 1. vyd. Bratislava : EI & T, 2002. 376 s.