
Rešeršní činnost

Vyhledávací nástroje internetu 1

18. 4. 2008, 2. 5. 2008

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

O co nám půjde?

- ❑ vyhledávací nástroje/služby, které slouží k vyhledávání informačních zdrojů dostupných volně/veřejně v prostoru www - **prohledávání tzv. povrchového webu**
- ❑ vyhledávací nástroje/služby **neviditelného webu** (*Invisible Web - Sherman*), **hluboký web** (*Deep Web - Bergman*) – informace přístupné přes www, které vyhledávací stroje nedokáží indexovat nebo z určitých důvodů nechtějí zahrnout do svých indexů

Základní typy

vyhledávací stroje (*search engines*)

- databáze vyhledávacího stroje je budována na základě automatizovaného sběru dat (*robot, worm, červ, spider pavouk*)
- vyhledávání na základě klíčových slov

předmětové adresáře/katalogový vyhledávací nástroj

(termín TDKIV) - *subject directories, subject services, portals ...*

- sběr informačních zdrojů ručně, jejich kategorizace – (spolu)pracovníci služby, inf. profesionálové a odborníci z oboru, tvůrci stránek
- zpřístupňování IZ prostřednictvím schémat předmětové kategorizace – prohlížení (*browsing*)

Základní typy

vyhledávací nástroje zaměřené na nejrůznější typy obsahů

- např. diskusní skupiny (newsgroups), elektronické konference (mailing lists), obrázky, video
- součástí základních kategorií nebo samostatné služby
- ☞ prolínání jednotlivých typů, stírání rozdílů
- základní kategorie mohou v sobě zahrnovat i další funkce, např. ty vztahující se k portálům

Webové portály

- ☞ sídlo, které poskytuje širokou škálu služeb a informací, často s možností jejich přizpůsobení uživateli podle osobních potřeb a zájmů
- ☞ *vyhledávání webových informačních zdrojů a dokumentů*, freemailová služba, denní zpravodajství, mapy, vyhledávání kontaktních informací na osoby i firmy, zábava, elektronický obchod (TDKIV) – **personalizace**
příklady: Yahoo!, Excite, MSN Search, Seznam
- ☞ **horizontální portály** - široce (obecně) zaměřené portály
- ☞ **vertikální portály** - zaměřené na specializované inf. zdroje

Posun k personalizovaným portálům – personal portal

- ☞ **iGoogle** – personalizace je založena na tzv. **Google Gadgets** – miniaplikacích, které umožňují výběr vlastních témat,
 - ☐ lze je libovolně kopírovat a vkládat do webových stránek
- ☞ **My Yahoo** – všechno na jedné stránce včetně počasí, emailu, seznamu televiz. pořadů, map, záložek, horoskopy ...

Neustálé změny

☞ vliv Webu 2.0 na služby vyhledávacích nástrojů:

- ❑ vznik nových typů služeb
(např. Wikia Search, Swicki - collaborative social search engine, vyhledávání založené na folksonomiích)
- ❑ nové obsahy vyhledávání (např. uživateli řízený obsah, podcast, blogy)
- ❑ modifikace stávajících nástrojů a jejich služeb

☞ Google koupil Youtube

☞ Yahoo získal Flickr a Del.icio.us

☞ koupí Microsoft Yahoo? Yahoo dostalo od Microsoftu třítydenní ultimátum

starší změny

- ☞ Yahoo! – původně předmětový katalog, dnes fulltextové vyhledávání, portál
- ☞ Teoma – koupen Ask
- ☞ AltaVista – připojen k databázi Yahoo!, neplatí mnoho původních pokročilých způsobů pro vyhledávání
 - Dead Search Engines

Kde změny sledovat

- ☞ [Search Engine Showdown](#) – novinky, recenze, charakteristika jednotlivých služeb
- ☞ blogy specialistů, např. [Karen Blakeman's Blog](#), [Marydee Ojala](#)
- ☞ [WebSearchGuide](#)
- ☞ [FreePint](#)
- ☞ [Search Engine Guide](#)

- ☞ Notess, G. R. [Search Engine Update](#) – pravidelná rubrika magazínu *Online* (plný text dostupný přes db LLIS)
- ☞ **průběžné rešerše v oborových db**
 - ☐ např. v db LLIS reš. dotaz: [\(Internet searching\) <in> Subject\(s\)](#)
[OR \(Internet search engines\) <in> Subject\(s\)](#)
→ využití služby ALERTS

Předmětové adresáře

- ☞ angl. ekv. - web directories, subject directories, subject services, portals

organizovaný a strukturovaný přístup
k informačním zdrojům na internetu

- ☞ **různá schémata předmětové kategorizace**
 - hierarchický princip, podřazení specifictějšího tématu obecnému, seskupování příbuzných témat

Předmětové adresáře

Výhody

- ➡ selektivní – hodnocení užitečnosti a vhodnosti
- ➡ do určité míry záruka kvality
- ➡ kategorizují/klasifikují
- ➡ snadno prohledatelné
- ➡ dobré pro obecné dotazy
- ➡ většinou i přímé vyhledávání (searchability)

Nevýhody/slabá místa

- ➡ malá databáze
- ➡ nemusí obsahovat stránky se specifickým obsahem
- ➡ méně vyhledávacích technik
- ➡ placené včlenění stránek může ovlivnit kvalitu
- ➡ tendence indexovat pouze hlavní/úvodní stránky
- ➡ někdy zdlouhavá navigace

Hlavní znaky

- 👉 **selektivnost** – stránky přidávané editory/navrhované uživateli, měly by být obsahově relevantní k dané kategorii, někdy placené zařazení stránek, zahrnutí spíše hlavních stránek – ne všech úrovních web. sídla
- 👉 **prohlížení - kategorizace** – různá kategorizační schémata, různé počty hierar. úrovní
- 👉 **vyhledávání** – vyhledávání v databázi adresáře, základní techniky, někdy i rozhraní pro pokročilé v.

Kdy využijeme adresáře?

- **hledání odpovědí na obecné otázky** – prohlížení dostupných zdrojů
 - **složená témata** – skládají se z více pojmů/dílčích témat – **vyhledávací stroje**
 - např. seznam vysokých škol z oblasti LIS X **výuka informační gramotnosti formou e-learningu na vys. školách**
- **hledáme sice specifickou informaci, ale nedokážeme ji vymezit** - potřebujeme informace ke specifikaci dotazu
- **získáme většinou odkazy na webová sídla k danému tématu** a ne seznam stránek obsažených v těchto sídlech

Komerčně orientované adresáře

- oslovení co nejširší veřejnosti
- oblasti společenského života – zábava, sport, obchod ...
- omezená anotace, různá kvalita
- Open Directory Project – 590 tis. kategorií, 4,5 mil. odkazů, 67 576 editorů
- Yahoo!Directory – 3 mil. odkazů

Oborové informační brány/předmětové/tematické brány

- podpora rozvoje výzkumu, vědy a vzdělávání
- zaměřeny oborově nebo univerzálně
- zhodnocené IZ
 - výběr na základě kritérií, informační specialisti, odborníci
- požadavek kvalitního předmětového přístupu
- více viz. např. Stoklasová, T.

KDY použijeme OIB?

- ➡ chceme-li získat odkazy na webové stránky, které doporučují, hodnotí a anotují experti
- ➡ zabýváme se o konkrétní obor, temat. oblast
- ➡ **nevyžadujeme mnoho na výstupu**
- ➡ chceme-li se vyhnout IZ s nízkým obsahem, které vracejí vyhledávací stroje

Co vědět?

- ☞ co hledáme na internetu, musíme znát předmět, o který se zajímáme
- ☞ do které předmětové kategorie patří hledaná informace
- ☞ musíme být ochotni věnovat čas navigování hierarchickou strukturou

OIB - *quality-controlled subject gateways/subject gateways*

významné OIB univerzálního charakteru:

Librarians' Internet Index – zdroje vybrány s ohledem na uživatele veřejných knihoven, anotované zdroje, popis pomocí LCSH

Internet Public Library - founded by a class at the University of Michigan's School of Information, and Michigan SI students

Infomine – LCSH, LCC

Academic Info

Intute (Resource Discovery Network) - zdroje vybrané více než 70ti tisíci vzděláv. a výzkum. institucemi, lze prohlížet 4 samostatné sbírky

Vyhledávací možnosti

Internet Subject Directories



Recommended Subject Directories



Specializované adresáře

- ☞ výborná výchozí místa pro poznání zdrojů z konkrétních oblastí
- ☞ kvalitní spec. adresáře - OIB
- ☞ **angl. ekv.** – specialized directories, resource guides, metasites, cyberguides, Webliographies, collections of links

rozmanitost:

- ☞ oborově nebo odvětvově orientované
- ☞ zaměřené na specifický druh dokumentu (noviny, historické dokumenty)
- ☞ vertikál. portály/vortály ([MEDLINEPlus](#))

Výhody

- ➡ specializace
- ➡ velmi selektivní
- ➡ poskytují okamžité odborné stránky

Nevýhody

- ➡ malé
- ➡ variují co do kvality a konzistence
- ➡ nemusí nabízet pole pro vyhledávání

Jak nalézt specializovaný adresář?

Přes adresáře:

- ☞ Yahoo - v rámci temat. kategorie podkategorie *Directories* či *Web D.*
- ☞ zadat klíčová slova do vyhledávacího boxu

Sledování odborného tisku, který se zabývá internetovými zdroji pro specifické oblasti

- ☞ časopisy **Online**, **Searcher**, webové stránky pro rešeršéry, např. [Free Pint](#)

Jak nalézt specializovaný adresář?

Užití vyhled. strojů

- ☞ volba vhodných vyhl. výrazů pro obor a adresář

Prostřednictvím stránek institucí

- ☞ vysoké školy, knihovny, výzkumné instituce, stránky zájm. sdružení

Užití adresářů adresářů (directories of directories), např.

- ☞ [WWW Virtual Library](#) - specializ. adresáře uspořádané do kategorií, tvoří dobrovolníci – sekce – různé formy
- ☞ [Search Engine Guide](#)

V čem se liší adresáře?

- ➡ velikost
- ➡ kategorizace
- ➡ anotace
- ➡ vyhledávání pomocí polí
- ➡ tvůrce
- ➡ rysy portálů

Příklady adresářů

Zaměřené na referenční informace

☞ refdesk.com

☞ [Internet Public Library Reference Ready Reference](#)

Polytematické zaměřené na akademickou a výzkumnou oblast

☞ [InfoMine](#)

☞ [BUBLLINK](#)

☞ [Intute \(Resource Discovery Network\)](#)

Příklady oborových adresářů LIS

- ☞ Informace pro knihovny
- ☞ Internet Library for Librarians
- ☞ Oborová brána Knihovnictví a informační věda

Neviditelný web

- ➡ rozmanitý obsah na internetu, který je „neviditelný“ pro vyhled. stroje
- ➡ odhady velikosti NW – 200 – 500krát větší než povrchový web
- ➡ obsahuje důležité zdroje
- ➡ existují způsoby, jak vyhledávat v NW

Proč vyhl. stroje neindexují obsah?

- ➡ informace uložené v databázích (až 54 %)
- ➡ stránky tzv. **samotáři**, z kterých není odkaz na jiné a na které není odkaz
- ➡ **dynamicky generované stránky** (po základní interakci s uživatelem), stránky vytvořené na základě uživatelského požadavku
- ➡ **stránky chráněné heslem, přístupné po registraci**
- ➡ robot navštíví sídlo, ale **hledá pouze do určité hloubky**
- ➡ autor použil tzv. **NO-ROBOT TAG**

“Pravý” neviditelný web

- ☞ materiál generovaný z databází, např. telefonní seznamy, patent. db, biograf. a plnotext. db
- ☞ často neexistuje unikátní URL
- ☞ stránka je vytvořena na základě uživateleova požadavku - roboti neumí zadávat dotazy

Brány pro neviditelný web

Příklady služeb, které registrují zdroje

neviditelného webu (v podstatě se jedná o adresáře databází, které jsou vytvářeny informačními profesionály. Mají navíc často přidanou hodnotu díky popisu jednotlivých zdrojů)

☞ Complete Planet - prohledává 89, 000 databází a speciálních vyhledávacích strojů, neobsahuje pouze neviditelné obsahy, ne vždy užitečné zdroje

☞ OIB, specializované adresáře – Infomine, Librarians' Index, BUBL Information Service

☺ některé vyhledávací stroje umožňují
vyhledávání informací na
viditelném i neviditelném webu

↓ ↓
informace relevantní pro vědu, výzkum
a vzdělávání:

SCIRUS

Google Scholar



specializovaný vyhledávací stroj

- ☞ zaměřuje se na odbornou literaturu, vyhledávání dokumentů z různých oborů a z různých IZ
- ☞ články, knihy, preprinty, kvalifikační práce
- ☞ IZ: vydavatelé odborné literatury, profesní společnosti, db preprintů, univerzitní a jiné odborné instituce, databáze



Každý výsledek vyhledávání reprezentuje skupina patřící k nějakému odbornému textu.

Může jít o jeden nebo více vztahujících se článků, více verzí jednoho článku.

Např. součástí výsledku vyhledávání může být skupina článků včetně preprintů, konferenčních příspěvků, časopiseckých článků, které jsou asociovány s jedním informačním dotazem.



Příklady:

- ☞ allintitle:"Models in information behaviour research" author:wilson
- ☞ allintitle:"Modern information retrieval" author:Baeza-Yates
- ☞ *vzdělávání dospělých v knihovnách se zřetelem na zlepšení jejich informační gramotnosti*
- ☞ **"adult education" library ("information literacy" OR "information skill")**



Odkazy u vyhledaného dokumentu

- ☞ **Cited By** – Identifies other papers that have cited articles in the group
- ☞ **Related Articles** – Finds other papers that are similar to articles in this group
- ☞ **Library Links (online)** – Locates an electronic version of the work through your affiliated library's resources . These links appear automatically if you're on campus.
- ☞ **Library Links (offline)** – Locates libraries which have a physical copy of the work.
- ☞ **Group of** – Finds other articles included in this group of scholarly works. Examples include preprints, abstracts, conference papers or other adaptations.
- ☞ **Web Search** – Searches for information about this work on Google.
- ☞ **BL Direct** – Purchase the full text of the article through the British Library.



Library Links

- ☞ Je možné získat přístup ke konečnému dokumentu prostřednictvím knihovny.
- ☞ Google spolupracuje s knihovnami, aby určil, které časopisy a texty jsou jimi předplacené elektronicky a potom spojuje články z tohoto zdroje, když jsou dostupné.
- ☞ Link se objeví, pokud se vyhledává z určité knihovny/informační instituce, automaticky se zahrnou tyto linky.



Řazení výsledků podle relevance:

- ☞ při setřídění výsledných odkazů bere kromě plného textu článku v úvahu autora článku, publikaci, ze které článek pochází, a jak často byl článek citován
- ☞ jako samostatné výsledky prezentuje také **citace**, i když nejsou citované dokumenty online dostupné

SCIRUS

for scientific information only

- ☞ specializovaný vyhledávací stroj firmy Elsevier pro sféru vědy, výzkumu a vzdělávání
- ☞ prohledává přes 250 miliónů odborně zaměřených webových stránek:
 - *identifikuje vědecké, odborné, technické a lékařské informace na www*
 - *nalezne nejnovější vědecko-výzkumné zprávy, články, patenty, preprinty, časopisy, které nenaleznou vyhledávací stroje*
 - *filtruje nerelevantní informace nepatřící do odborné sféry*

SCIRUS

for scientific information only

indexuje webové zdroje z viditelného webu

- 83 million .edu sites
- 25 million .org sites
- 10 million .ac.uk sites
- 22 million .com sites
- 6.5 million .gov sites
- over 68 million other relevant STM and University sites from around the world

indexuje zdroje z placených i volně dostupných bází dat

SCIRUS

for scientific information only

Řazení výsledků podle relevance:

- ☞ *Scirus vypočítává relevanci podle frekvence vyhledávaného termínu a počtu odkazů na stránku. Nabízí také možnost řazení podle data.*

Povinná literatura

- ☞ Hock, R. The extreme searcher's Internet handbook : a guide for the serious searcher. 2nd ed. Medford, N.J. : CyberAge Books, 2007. **Kap. 2 General Web Directories and Portals, 3 Specialized Directories, s. 29-61.**
- ☞ Makulová, S. *Vyhľadávanie informácií v internete : problémy, východiská, postupy*. 1. vyd. Bratislava : EI & T, 2002. **Kap. XV. Neviditeľný alebo hĺbkový web, s. 285-299.**

Prostudovat nápovědu:

- ☞ **Google Scholar Help + Advanced Search Tips** [online]. [cit. 2008-04-27]. Dostupné na World Wide Web: <http://scholar.google.com/intl/en/scholar/help.html>
- ☞ **SCIRUS help : Search tips Introduction** [online]. [cit. 2008-04-27]. Dostupné na World Wide Web: <http://www.scirus.com/html/help/index.htm>

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.
- ☞ Hock, R. **Search Engines : From Web 0.0 to Web 2.0 and Beyond**. *Online*, 2007, roč. 31, č. 1, s. 26-30.
- ☞ Makulová, S. **Vyhľadávanie informácií v internete : problémy, východiská, postupy**. 1. vyd. Bratislava : EI & T, 2002. 376 s.
- ☞ 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>