



FAKULTA SOCIÁLNÍCH VĚD
Univerzita Karlova v Praze

Vyhledávání na internetu

Michala Sošková
knihovna FSV



FAKULTA SOCIÁLNÍCH VĚD
Univerzita Karlova v Praze

jak fungují vyhledávače
google
ti druzí
neviditelný web
trendy



jak fungují vyhledávače

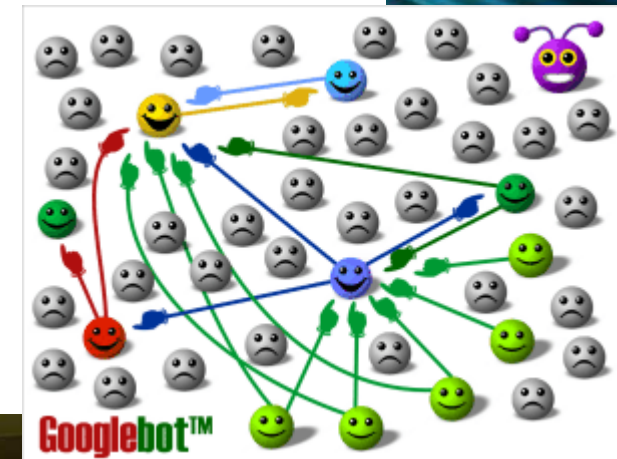
/ robot – spider/crawler

/ index



robot

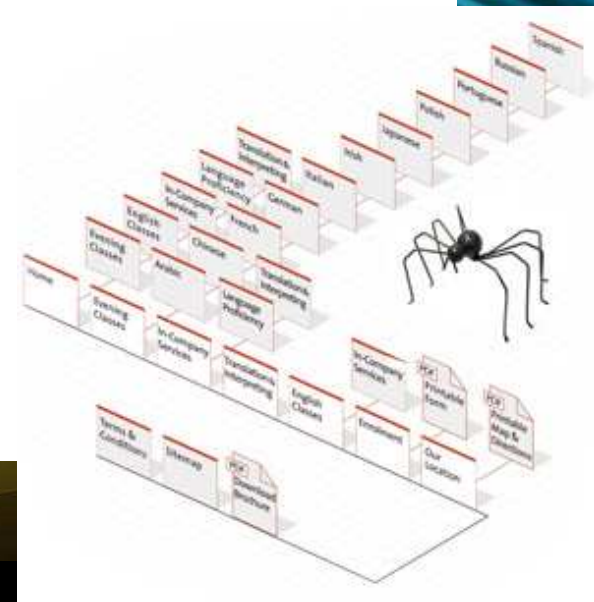
- / prochází stránky
- / stáhne klíčová slova + informace o nich
- / => seznam všech slov = **index**
- / zapamatuje si adresy > „mapa webu“
- / vrací se zpět





index

- / robot současně vytváří „mapu webu“ = index = seznam všech slov
- / při zadání dotazu vyhledávač neprohledává všechny stránky /to by trvalo velmi dlouho/, ale prohledává **index**





index

„jak správně vyhledávat na internetu“

1 internet-u (webu)

URL1, **URL2**, URL3

~~jak~~

~~na~~

2 správn- ě (-ý, -á, -í... vhodný, přesný...) **URL2**, URL4, URL5

3 vyhled-ávat (-at, -ám...) URL6, **URL2**, URL3



index

řazení výsledků ovlivňuje:

URL

titulek

podnadpis

četnost slov

začátek vět

.....



FAKULTA SOCIÁLNÍCH VĚD
Univerzita Karlova v Praze

a pak přišel google...



google

- / je tu 12 let
- / 2 doktorandi Kalifornské univerzity
- / vycházeli z akademické sféry, citovanost odborných prací

- / **PAGE RANK = RELEVANCE!!!**
- / **1-10 (10 má google)**

- / googol - číslo, které lze napsat jako jedničku se sto nulami > překlep google



a kde jsou peníze?

- / **NEAGRESIVNÍ REKLAMA** – pro každého
- / **ŠIROKÉ PORTFOLIO** – e-mail, kancelářské aplikace, fotky, zprávy, prohlížeč, operační systém...
- / **část vždy zdarma**



google

<http://www.youtube.com/watch?v=BNHR6IQJGZs>



vyhledávání na googlu

- / google.com x google.cz (prohledává přednostně domény.cz)
- / př: singapur
- / zpřesňování
- / vyhledávání různých formátů
 - obrázky
 - př. magnetismus
 - dle souboru
 - videa



vyhledávání na googlu

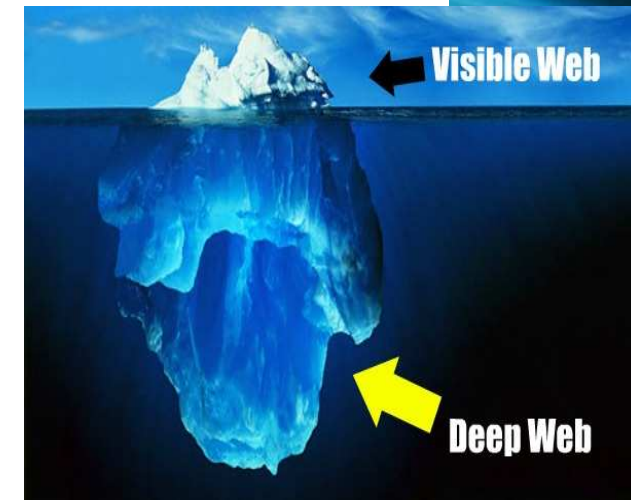
/ rozšířené vyhledávání

- novinář + žurnalista
- novinář -žurnalista
- novinář OR žurnalista
- „novinář“



neviditelný web

- / neprolinkované stránky
- / placené databáze x nepřístupné databáze
- / zaheslované stránky (is)
- / netextové informace (obrázky, videa, software... bez popisků)
- / dynamicky tvořené stránky (java script - letenky)
- / sociální sítě (?)
- / zakázané autorem
- / změněné stránky – webarchiv.cz, archive.org





google books

/ books.google.cz (2004)

- 129,864,880 (srpen 2010) mil. knih na světě,
- knihy s volným právem užití, ke stažení zdarma v pdf (dle licencí)
- web 2.0 funkce
- vyhledávání uvnitř knihy



příklad

- McLuhan – plný text



google scholar

/ scholar.google.cz

/ Google Scholar je specializovaný vyhledávač odborných textů (např. články z časopisů, preprinty, vysokoškolské kvalifikační práce). ScholarSFX je speciální verze link serveru SFX, která slouží k propojení výsledků vyhledávání v Google Scholar s plnými texty v databázích (např. EBSCO, Proquest, JSTOR)

/ citovanost, podobné články, verze

/ plné texty

/ vyhledávají i v některých licencovaných databázích, lze tedy zjistit, kde se článek nachází a objednat přes MVS

/ odkaz na knihu, pokud je ve fondu knihovny, kterou navštěvujete –
Předvolby - ne vždy funkční



ti druzí???

seznam.cz

bing.com

metavyhledávače

<http://www.dogpile.com/> (google, bing, yahoo, ask)

<http://www.metacrawler.com/>

vizuální vyhledávače

<http://search-cube.com>

<http://spezify.com>

<http://taggalaxy.de>



portály, rozcestníky (katalogy)

- / rozcestník / katalog / brána
 - / strukturovaný obsah, kategorie
 - / když nevíte, kde začít
 - / <http://www.caramba.cz/>
 - / <http://www.dmoz.org/>

- / portál
 - / komplexní služby
 - / vyhledávání + katalog + zprávy + e-mail
 - / seznam.cz



současné trendy

/ sociální sítě

/ 10 SOCIAL NETWORK SEARCH ENGINES FOR JOURNALISTS

<http://davidhiggerson.wordpress.com/2011/07/05/10-social-network-search-engines-for-journalists/>

/ <http://www.socialmention.com>

/ <http://search.twitter.com/>

/ <http://youopenbook.org>

/ <http://www.klaboseni.cz>

/ personalizace

/ spojování

/ rychlost, co nejmenší nároky na uživatele

/ google myslí za nás...



zajímavosti

/ <http://www.googlezeitgeist.com/cs>