

Vyhledávání na internetu

Východiska

Vyhledávací služby

Východiska

Na úvod je nutno si uvědomit, že internet nikdo nevlastní, nikdo jej neřídí a není v něm žádná centrální autorita, která by byla povinna vědět o všem a dokonce ani mohla vědět o všem. Výjimkou je centrální koordinace přidělovaných IP adres.

□ Otázkou je: **Je to dobře nebo špatně?**

😊 **dobře** – podporuje to skutečnou demokracii

😞 **špatně** – na internet se dostane i to, co by tam být nemuselo (nemělo).

Na internetu není žádné povolovací řízení. Když chce někdo něco zpřístupnit, nepotřebuje k tomu žádné povolení. Žádná autorita, která by věděla o všem tedy není, jsou jen pokusy vést evidenci dostupných zdrojů.

□ Otázkou je: **Jak posoudit věrohodnost informací na internetu?**

😞 **Těžko**, není žádná autorita, je jen nutno posuzovat věrohodnost zdroje informací.

Internet není **knihovna**, která má systém evidence knih a jiných informačních zdrojů, internet je mnohem více chaotický, nemá jednotný formát dostupných informací.

Pokud má být **vysoký stupeň organizace**:

- musí informace posuzovat **člověk**, což je pomalé a drahé
- bude **menší objem evidovaných informací** (což plyne z předchozího).

Pokud mají být zpracovávány **velké objemy dat**:

- dokáže to jen **stroje – počítače**
- stroje nedokážou posoudit dobře „o čem to je“, „jak a k čemu je to dobré“...

□ Otázkou je: **Jaký je princip vyhledávací služby?:**

- **vyhledávací služba** systematicky shromažďuje informace o tom, **co kde je k dispozici**
- **uživatelé**, kteří něco hledají, **se obracejí na vyhledávací službu** a ne na místo, kde tuší, že by se informace mohla vyskytovat.
- **vyhledávací služba** poskytne „ukazatel“ na **výskyt** hledané informace.

Vyhledávací služby

Rozdělení vyhledávacích služeb:

- **specializované**
- **univerzální**
 - *seznamy zdrojů*
 - „*knihovny*“ (*vyhledávací služby*)
 - *katalogy*
 - *vyhledávače (fulltextové vyhledávací služby)*
 - *vyhledávací rozhraní*
 - *metavyhledávače*

Specializované vyhledávací služby

Samostatné vyhledávací služby se dnes už moc nepoužívají, protože jsou specializované na určité vyhledávání a mají specializované servery a specifické klientské programy. Patří sem např.:

Informatika a výpočetní technika

- *Archie, Nosey Parker* – vyhledávání souborů
- *WAIS* – vyhledávání v dokumentech
- *WHOIS* – vyhledávání lidí
- *Veronica* – univerzální vyhledávací služba pro gopher.

Uživatelé dnes dávají přednost službám, které běží na www, neboť jsou jednoduché a mají univerzální ovládání. Proto se některé specializované služby integrovaly do www nebo vznikly rovnou jejich moderní nástupnické verze přímo v prostředí www.

Univerzální vyhledávací služby

Seznamy

Seznamy jsou zaměřeny na určitou úzkou oblast a obsahují jen malou část z toho, co je na internetu dostupné. Tvoří je většinou nadšenci jako svého koníčka, zahrnují velký objem práce a výsledek závisí na schopnostech autora.

Př.: Materiály o internetu: <http://home.pf.jcu.cz/pepe/>

„Knihovny“

Do určité míry jsou podobné seznamům, protože je tvoří lidé. Jsou však výsledkem práce nikoli jednotlivců, ale celých týmů pracovníků, kteří prohledávají internet a vyhledávají kvalitní zdroje informací. Říká se jim „knihovny“, protože jejich organizace knihovny připomíná, či dokonce nějaká významná knihovna za nimi přímo stojí. Obsahují kvalitní zdroje, často jsou to celé encyklopedie, které kromě vlastních článků jednotlivých hesel obsahují i řadu kvalitních a prověřovaných odkazů do celého internetu.

Tvůrce tohoto textu nezná žádnou českou knihovnu, zná pouze zahraniční, které jsou většinou v angličtině. Pouze základní všeobecné informace na „knihovnách“ bývají zdarma, podrobnosti pak už většinou za poplatek. Jsou vhodné pro získání skutečně kvalitních a obsažných materiálů.

Př.: <http://www.britannica.com>

Katalogy

Jsou zaměřeny univerzálně. Jejich tvůrce vytváří předmětovou hierarchii sekcí, do které jsou zařazovány odkazy. Katalogy vyžadují velký objem lidské práce, jsou proto většinou výsledkem profesionální práce týmů tvůrců. Uživatel si prochází soustavu nabídkových stránek a vyhledává si odkazy, které ho zajímají. Sekce na nejvyšší úrovni vytváří provozovatel katalogu, zařazování odkazů do sekcí provádějí uživatelé sami (po následném schválení a definitivním zařazení správcem katalogu) či dokonce uživatelé spoluvytvářejí i podsekce uvnitř sekcí.

Fulltextové vyhledávače

Narozdíl od katalogů neobsahují žádné sekce, pouze se snaží „mapovat“ co nejvíce internetových stránek a co nejvíc jich poskytnout uživateli, který si zadal vyhledávací dotaz.

V internetu neexistuje jediná „ta pravá“ vyhledávací služba a ani žádná univerzální vyhledávací strategie. Existuje mnoho vyhledávacích služeb, ale každá je v něčem trochu jiná. Platí obecně, že pokud uživatel informaci nenašel, neznamená to, že se na internetu nevyskytuje. Takže má smysl zkoušet i další vyhledávací služby. (To platí nejen pro fulltexty, ale i pro katalogy.)

Fulltext versus katalog

Fulltext je pro uživatele, který:

- dost přesně ví, co chce a umí dobře to dobře popsat – tj. sestavit vyhledávací dotaz
- potřebuje prohledat co nejširší okruh zdrojů

Katalog je pro uživatele, který:

- nemá konkrétní představu, ví jen, že informace patří do určitého oboru poznání

- nehledá nic konkrétního, ale chce objevovat nové věci, které by ho mohly zaujmout
- se chce cíleným způsobem (tj. zaměřeným na určitou problematiku) „toulat“ (surfovat) internetem.

Katalog + fulltext = portál

Moderní vyhledávání razí zásadu:

„Uživatel by nikdy neměl odejít s nepořízenou“, tedy vždy nabídnout alespoň něco. Katalogy a fulltexty se kombinují a vznikají **portály**, které se již nesnaží uživatele nasměrovat dál (jako katalog), ale zadržet a nepustit dál, poskytnout mu všechno, aby neměl potřebu odcházet jinač.

Portál je soubor služeb zaměřených na široké spektrum uživatelů. Zahrnuje obvykle:

- katalog
- fulltextový vyhledávač
- freemail (el. poštu zdarma)
- freeweb (prostor pro vlastní stránky uživatele)
- diskuse (chat, konference)
- zpravodajství
- specializované vyhledávání (lidí, v obchodech na internetu, v mapách aj.)

Příklady českých portálů:

- Seznam – <http://www.seznam.cz>
- Atlas – <http://www.atlas.cz>
- Centrum – <http://www.centrum.cz>

Příklady českých komunikujících fulltextových vyhledávačů:

- Google – <http://www.google.com> - zahraniční, umí hledat české texty, ale „česky neumí“
- Juxo – <http://www.juxo.cz> - český, plně podporuje češtinu (skloňování, časování aj.)
- Morfeo – <http://www.morfeo.cz> - český, plně podporuje češtinu

Kladení dotazů fulltextovým vyhledávačům

Podrobné informace jsou v návodech jednotlivých vyhledávačů. **Obecně platí:**

- ❖ Dotaz se zadává **klíčovým slovem**, popř. klíčovými slovy.
- ❖ Pokud je více klíčových slov, musíme určit **vztah mezi klíčovými slovy**
 - všechna klíčová slova
 - alespoň jedno klíčové slovo
 - přesně tak jak je napsáno
 - klíčová slova „blízko sebe“
- ❖ přesně tak jak je napsáno – **fráze** – se dává do uvozovek např. „to be or not to be“
- ❖ **booleovské vyhledávání** – užívá se booleovských (logických) operátorů:
 - AND – obě slova současně
 - OR - alespoň jedno slovo
 - NOT – slovo se nesmí vyskytovat
 - NEAR – slova blízko sebe (např. maximálně do vzdálenosti 10 slov)
- ❖ někdy se **booleovské operátory nahrazují**
 - „přirozenou řečí“ – mezi dvě slova vyhledávač automaticky dosazuje tzv. měkký AND (zprvu ho respektuje, v dalších odkazech už hledá jen jednotlivá slova)
 - znakem + (slovo se určitě vyskytovat musí např. +Praha)
 - znakem – (slovo se určitě vyskytovat nesmí)

Vyhledávací rozhraní

Jsou to formuláře, které umožňují klást dotazy různým vyhledávacím službám z jednoho místa. Odpověď pak vrací přímo příslušná volaná služba. Hlavní výhodou je pohodlí uživatele, který nemusí znát odkazy na jednotlivé vyhledávací služby.

Př. vyhledávací služba Alenka - <http://www.alenka.cz>

Metavyhledávače

Jsou vyhledávače, které přijmou jednotlivé dotazy a tyto dotazy předají vyhledávacím službám podle plánu vyhledávání, který si samy sestaví. Počkají na odpovědi a pak je vracejí uživateli uspořádané do jednoho celku (bez duplicit).

Př. světových metavyhledávačů (české autor nezná – nejspíš žádné nejsou)

- DogPile – <http://www.dogpile.com>
- MetaSearch – <http://www.metasearch.com>

Další světové vyhledávače

- Altavista – <http://www.altavista.com> - portál
- Yahoo! Search – <http://www.yahoo.com> - portál
- AllTheWeb – <http://www.alltheweb.com> - fulltextový vyhledávač

Některé příklady speciálního vyhledávání na www

- lidí
- souborů
- domén
- obrázků
- multimediálních souborů
- odkazů na konkrétní stránky
- v mapách
- dopravního spojení
- v internetových obchodech

Literatura:

[1] Archiv článků Jiří Peterky

[2] různé zdroje na www