

A keresés eszközei

Keresés fajtái

- Matching (szabadszavas)
- Browsing (böngészés, „nyomkövetés”)

Internet felépítése

- Szöveges (lynx)
- Grafikus (WWW)

World Wide Web felépítése

- Hipertext
- Multimédia
 - képek

Internetes keresés buktatói

- Halott linkek
- „szürke irodalom”
- Irreleváns találatok (pl. „library”)
- Minden kereső más-más részterületet pásztáz, sosem az egészet
- Nem indexelt oldalak nem kerülnek elő
- Adatbázisokat nem indexelnek
- Nyelvi problémák:
 - Gépi nyelv logikája (nem alkalmas szabadszavas keresésre)
 - Idegen nyelvtudás hiánya

Megoldások

- Hierarchikus katalógus
(*kifejezetten könyvtáros szakterület*)

Keresőrobotok

- Automatikusan feltérképezik a web hipertext szerkezetét
- Összeszedik a keresésnek megfelelő dokumentumokat

Ágensek

- Böngészőprogramokba beépítettek
- Valamilyen algoritmus szerint fontossági sorrendbe rendeznek, szűrnék: web-tartalmat, leveleket.
- Felhasználói igények minél pontosabb felmérése

Ágensek

- Önálló ágensek
 - Site-ok közt járkálnak, önállóan döntenek el, mikor hagyják el az oldalt és mi a teendőjük. Csak speciális szerverek közt jelennek meg.

Ágensek

- Intelligens ágensek
 - Pl. termék kiválasztásában segítenek a felhasználónak. Vagy egy űrlap kitöltésében. Vagy dolgok megtalálásában. Nem sok közülük van a hálózathoz.

Ágensek

- Felhasználói ágensek
 - Pl. web-felhasználói ágensek: „What's related”
 - NNavigátor
 - MS IE
 - E-mail felhasználói ágensek:
 - pl. Eudora

Kereső motorok

- Bizonyos adatkészleteken keresztül keresnek.
- Webes értelemben: HTML-dokumentumok olyan adatbázisában keresnek, aminek tartalmát robotok gyűjtötték össze.

Robotok felhasználása

- indexelésre
- HTML érvényesítésre
- Link érvényesítésre
- "What's New" figyelésre
- tükrözésre

Indexelő robot

Honnan tudja, hová kell menni?

- Eltérő stratégiák:
 - URL-ek history-listáiból
 - **Szerver listák**
 - „What's New”
 - Legnépszerűbb oldalak
 - kézzel is bővíthető URL-lista.
 - ← USENET-levelek, publikus levelezőlista-archívumok
 - ← kiindulópontok

Indexelő robot 2.

Honnan tudja, mit indexeljen?

- Elemzés tárgya:
 - HTML-cím
 - Bekezdés első sora
 - Az oldal tartalma
 - META-tag
 - Más rejtett tartalom
- Minden szó indexelése vagy rangsorolása

Katalógusfajták

- Hierarchikus katalógus
- Tárgyszó-katalógus
- Tezaurusz
- ETO
- Dewey-rendszer

Módszerek

1. Szabadszavas információkeresés
2. Kötött szavas információkeresés
 1. Keresőszó megfogalmazása
 2. Böngészés
 3. Véglleges keresőszó
 4. Célzott keresés
 5. Találatok elemzése