

(22) Internetes keresőrendszerek használata

Az internet ma már nem csupán kapcsolattartási lehetőséget biztosít, hanem számottevő mennyiségű információ is felhalmozódott a hálózaton. Ezen információk többsége szabadon elérhető. Jelentős probléma, hogy az ebben az adathalmazban való keresés nem egyszerű feladat egyrészt a mennyiség, másrészt a központi katalogizálás hiánya miatt. (Utóbbi az internet fejlődéséből egyenesen következik.)

Beláthatatlan adatmennyiség van a weben

- A webhelyek, portálok, adatbázisok ugyan rendezettek, de a web egésze rendezetlen, gyorsan fejlődő, változó rendszer-> egyszerű böngészéssel, webhelyek és azonosítók (URL) alapján nehéz valamit megtalálni.
- Az interneten számos olyan speciális **keresőgépet** hoztak létre, amelyek igyekeznek a web minél nagyobb részét feltérképezni, és a talált lapok mutatóit, adatait tárolják. Olyan keresőprogramot futtatnak, amelyet a webhely látogatói saját böngészőjükből vezérelhetnek.

A webes tartalmak megjelenésétől igényként merült fel a felhalmozott információk közötti eligazodás megkönnyítése. Ezért régóta fontosak a kereső szolgáltatások. Két módját szokás megkülönböztetni: a tematikus (témakör szerinti) és a kulcsszavas (tartalom szerinti) keresést.

Kulcsszó: a keresett adatra, információra jellemző szó, illetve szöveg. A kulcsszavak alapján történő keresést **kulcsszavas keresésnek** nevezzük.

Internetkatalógus: a kereső webhelyen kialakított, többszintű, hierarchikusan rendezett tematikus rendszer. Emberi munkával rendezik a weblapok adatait. Ebben téma szerint kereshetünk egyre szűkülő körben, amit tematikus keresésnek nevezünk.

Általános információs oldalak: pl. www.lap.hu, www.kereso.lap.hu, -> speciális keresők.

A kulcsszavas (keresőgépes) szolgáltatás

Keresőgépek: Google (www.google.co.hu), AltaVista (www.altavista.matav.hu), Excite, Infoseek, AltaVista.

Kereshetünk oldalakat (dokumentumokat) egyszerű kulcsszóval vagy kulcsszavakból alkotott kereső kifejezés segítségével

- a gép kiírja a keresés eredményét (a találatok számát, a keresési időt) és a találati lista első néhány elemét,
- kiírja a talált dokumentumok néhány adatát: a dokumentum címét, a dokumentum webcímét (URL-jét), terjedelmét (adatmennyiségét), a megjelenés dátumát stb.

Egyszerű keresés: Egy tárgyszavas kereső webhelyen adjuk meg a kérdéses kulcsszót. Érdemes - amennyiben magyar szóra keresünk - a ragozatlan alakkal is próbát tenni. Ha nem elegendő a találatok száma, vagy egyáltalán nincs találat, akkor finomítsuk a kereső kifejezést. Rokonsértelmű vagy azonos alakú kifejezések is okozhatnak problémát.

Összetett keresés: A találat sokkal pontosabb lehet, ha nem egyetlen szóval, hanem szavak egy sorozatával végezzük el. Fontos a szavak sorrendje és a ragozott vagy ragozatlan alak is, ez megegyezik az egyszerű kereséssel.

Kereső kifejezések

Tovább finomítható a keresés, ha a több kulcsszóból álló füzérben speciális karakterekkel adjuk meg a keresési feltételeket.

A szavakból műveleti jelek (operátorok) segítségével kifejezések képezhetők

- a * tetszőleges karaktereket helyettesít (töredékszó használata), például: *játék** minden olyan szó, amely a „játék” szóval kezdődik,

- „karakterek láncolata” pontosan a megadott karakterláncra, kifejezésre keres. ez lehet több szó, használhatunk kis- és nagybetűket,
- NOT - az utána álló szó nem fordulhat elő a dokumentumban (tagadás művelete),
- a +jel után álló szónak minden dokumentumban szerepelnie kell,
- a - 'jel után álló szó egyik dokumentumban sem szerepelhet,
- NEAR - az előtte és utána álló szavak legfeljebb 10 szónyi távolságra lehetnek egymástól a dokumentumban,
- AND - az előtte és utána álló mindkét szónak elő kel! Fordulnia (ÉS művelet),
- OR - az előtte és utána álló szavakból legalább az egyiknek elő kell fordulnia (VAGY művelet).

Szűrők használata - korlátozó, szűkítő feltételek

- a dátum szerinti szűrés a dokumentum megjelenési időintervallumát korlátozza,
- az ékezet nélküli szűrés nem veszi figyelembe az esetleges ékezeteket,
- a nyelv szerinti szűrés csak az adott nyelvű dokumentumokat veszi figyelembe,
- csak a webcímben (URL) keres,
- csak a dokumentum címében (title) keres.

A kulcsszavas kereső általában sok találatot eredményez. A találati lista relevancia szerinti, azaz a keresőszerver szerint lényegi tulajdonságok dolgában a keresettel legjobban egyező találatok a lista elejére kerülnek. Kicsi tehát az esély, hogy a találati lista hátrébb levő eleme számunkra érdekes lenne, de a keresőszerver ilyen döntése nem mindig helyénvaló. A kulcsszavas keresés jól megfogalmazott kulcsszót vagy összetett kifejezést igényel és a találati lista áttekintése mindenképpen időigényes.

A részletes keresést gyakran űrlappal támogatják, ezt viszonylag könnyű kitölteni.

A kulcsszavas keresés során keresőhelyeket használunk. Ezek olyan webhelyek, melyek keresőmotorja saját adatbázist tart fenn. Ezekben egy-egy kulcsszóhoz egy vagy több URL tartozik. Ezt az adatbázist a webhely keresőmotorja időről időre frissíti. Egy weblap készítése során ritka kivételtől eltekintve fontos annak figyelembe vétele is, hogy az esztétikus megjelenés, a rendszerfüggetlen olvashatóság mellett kiszolgáljuk a keresőket is. Ezek ugyanis időről időre felkeresik a webszervert is futtató gépeket és a lehetőségeknek megfelelően az ezeken közzétett információkat katalogizálják. A weblap kialakításával segíteni lehet ezt, aminek pozitív hozadéka számunkra az, hogy lapunkat a kérdéses keresőt használva könnyebben találják meg majd ismerőseink, illetve kereskedelmi célú webhely esetén reménybeli ügyfeleink. Ehhez a HTML-nyelv is ad lehetőséget, többet is. A keresők előnyben részesítik a lapjainkon a HTML-nyelv *header* címkéit: a <H1>, <H2> stb. címkék közé helyezett szöveg katalogizálása prioritással bír a kereső számára. A HTML-állomány fejléc része - a <HEAD> címkepár között - lehet úgynevezett <META> címkét elhelyezni. Ennek keywords opciójú tagja lehetővé teszi kulcsszavak elhelyezését, sőt azt is jelezhetjük, hogy mely nyelven értelmezett ez a kulcsszó. A weblap szövege csak ezek után képezi a keresőmotorok érdeklődésének tárgyát, elsődleges tehát, hogy a fejlécben kulcsszavakat képezzünk és a szövegben szabályosan helyezzünk el különböző szintű címsorokat.

A keresés algoritmus

Kulcsszavak meghatározása ⇒ kereső kifejezés, szűrők ⇒ a keresés elindítása ⇒ a találati lista értékelése ⇒ ha az eredmény nem megfelelő, vissza az elejére.

Egyéb keresők

A metakeresők szétküldik a kérdést különböző keresőknek, és a bejött adatokat rendezik, a találatokat közös iistán jelenítik meg:

www.ariadnet.hu, www.mamma.com.

A speciális gyűjtőkörű keresők egy jól körül határolt területen keresnek (földrajzi területen vagy dokumentumtípus szerint, zenei fájlokat, képeket stb.).

Tematikus keresők - internetkatalógusok

A téma szerinti keresés pontosabb eredményt hoz, de a találatok száma kevesebb.

Keresőhelyek: Hudir (www.hudir.hu), Yahoo (www.yahoo.com), Lycos, Magellan.

A kereséssel kapcsolatban témaköröket ismerünk. Olyan webhely felkeresésével kezdjük, amely tematikusan csoportosított linkgyűjteményeket tartalmaz. Ezek üzemeltetői hirdetések megjelenítésével finanszírozzák vállalkozásukat, az oldalt meglátogatók számára a szolgáltatás ezért ingyenes. Magyar nyelven az egyik legteljesebb ilyen gyűjtemény a lap.hu címről érhető el.

Ez a lap többszintű. A nyitólap valamely linkje még nem egy lehetséges találati lapra, hanem szinte mindig egy második URL-gyűjteményre mutat, ahol az első lapon kiválasztott címmel kapcsolatos linkeket gyűjtötték össze.

A tematikus keresés előnye, hogy pontosabb, kevesebb a hamis találat. Sajnos az értékes is: téma szerint keresve kevesebb találatból válogathatunk, mint kulcsszavas keresés esetén.

A dokumentumok adatai tartalmuk alapján vannak csoportosítva. A főcsoportokban (kategóriákban) minden téma további altémákat tartalmaz mindaddig, amíg el nem érünk a dokumentumok szintjére