

Kaszab István: Információkeresés az Interneten

Az Internet jellegéből eredően az Interneten nincs központi katalógus, és a böngészőprogramok csak a pontosan megadott webcím felkeresésére képesek, nem rendelkeznek intelligens tulajdonságokkal. Így a számunkra ismeretlen helyen lévő weboldalak eléréséhez keresőprogramok vagy linkgyűjtemények, portálok segítségével van szükségünk.

A keresőprogram egy olyan szolgáltatás, mellyel kulcsszavakat használva az Interneten elérhető honlapok címére kereshetünk. Az eredményt a keresőoldalak rendszerint fontossági szempontok szerint rendezve jeleníti meg. A találatokat ennek ellenére kritikával kell fogadni, mert a találati lista elején többnyire fizetett hivatkozások állnak, s nem ritka, hogy egyes hivatkozások már nem létező honlapokra utalnak.

A keresőoldalak különböznek egymástól keresési módszereikben, de leginkább a találati listák rendezésében. A keresőprogramok két fő csoportját különböztetjük meg:

- a szabadszavas keresők
- a kategória szerint keresők, portálok.

A szabadszavas keresőoldalak

A szabadszavas keresésnél egy vagy több keresőkifejezést kell megadni.

Az, hogy keresésünk mennyire hatékony, sikerül-e a legmegfelelőbb hivatkozásokat megtalálni, nemcsak a keresőoldal képességein múlik, hanem azon is, hogyan választjuk meg a keresőkifejezéseinket.

Ahhoz, hogy képesek legyünk jól megválogatni keresőszavainkat, egy kicsit jobban meg kell ismernünk a keresőoldalak, más néven keresőmotorok működését.

A keresőmotorok (angol nevén *Search Engine*) olyan nagyteljesítményű számítógépek, melyek ún. keresőrobotok (*search robot*) vagy keresőpókok (*spider*) segítségével felkeresik az Internet honlapjait, és hivatkozások után kutatnak, majd indexelik azokat saját maguk számára. Az így létrehozott adatbázisokban keres a keresőoldal az általunk megadott keresőszavak alapján.

Mit lát a keresőmotor a honlapokból?

Ez nagyon fontos kérdés, mert ettől függ, hogy hogyan indexeli a keresőrobot a honlapokat, és hogy mely keresőszavakra milyen találatokat mutat nekünk.

A keresőrobot megvizsgálja a honlap HTML kódját és a szövegtörzsben szereplő szavakat, hivatkozásokat (angolul *link*). A szövegtörzsben talált szavak gyakorisága mellett az ún. Meta adatokat is számba véve meghatározza az oldal tartalmát.

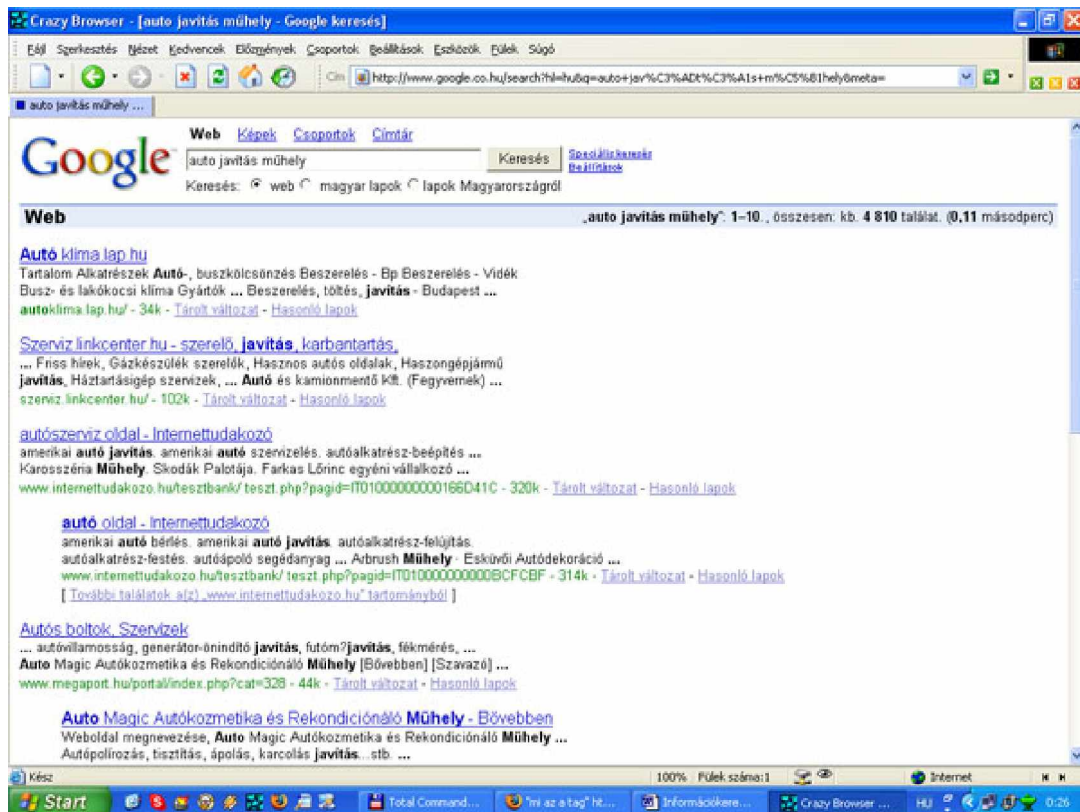
A Meta adatok a HTML kód fejrészében szerepelnek és a keresőrobotok munkáját hivatottak megkönnyíteni az oldal tartalmi besorolásánál.

Ezeket a Meta adatokat (angolul *Meta tag*) a honlap készítője készíti, ám az informatikus szakemberek képzettsége nem követel katalógizálási (könyvtárosi) ismereteket,

így csak részben használható. Túl gyakran fordulna elő, hogy a Meta adatok megtévesztenék a keresőmotorokat, esetleg hiányosak, vagy egyáltalán nincsenek is kitöltve. Ez az oka, hogy a keresőmotorok átvizsgálják a honlap szövegtörzsét is.

A keresés

Ilyen ismeretek birtokában, más sokkal könnyebb kitalálni, milyen keresőkifejezéseket használjunk a szabadszavas keresőkben. Nincs más dolgunk, mint meghatározni azt a tárgykört amit keresni kívánunk, és a témára legjellemzőbb kifejezéseket begépelni a kereső ♣



Meglehet, sőt, igen gyakori, hogy a számunkra legjobb találat nem az első sorokban van, gyakran nem is az első oldalon. (Mivel gyakran előfordul, hogy a találatok száma ezres nagyságrendekben mérhető, ezért a keresőoldalak többsége – az áttekinthetőség végett – 10 találatonként, oldalakra bontva jeleníti meg a találatokat.) Kis tapasztalattal gyorsan beletanulunk, hogyan válogassunk a találatok címhivatkozásait követő két-három soros leírásaiból, ami általában a keresőszót tartalmazó kiemelt szövegrészlet, vagy Meta adat idézet.

Sokszor ez a találati lista nem hoz kielégítő eredményt számunkra. Gyakran a találatok listájának túl sok oldalát kellene a átböngésznünk. Ennek elkerülésére alkalmazhatjuk a keresők többségénél használható Bool algebra műveleteit (George Boole (1815-1864) angol matematikusról). A keresőkben alkalmazható Bool műveletek használata jóval egyszerűbb, mint amilyennek hangzik. A Bool algebrai műveleteket meghatározó jeleket logikai operátoroknak, esetleg ÉS/NEM operátoroknak szokták nevezni. A logikai operátorokkal történő keresés nem más, mint olyan logikai előjelek alkalmazása, amelyek segítségével a találatok halmaza leszűkíthető a számunkra legfontosabb hivatkozásokra.

A keresőkben használatos logikai operátorok:

ÉS	– (AND) –	jele:	+
pontos kifejezés		jele:	” ”
NEM	– (NOT) –	jele:	–

A keresők többsége a matematikai jeleket (+/-) használja, ellenkező esetben erre felhívják figyelmünket a kereső honlapján.

A keresők figyelmen kívül hagyják a kis és nagy betűket, tehát mindegy, hogy „Budapest”-et, „budapest”-et vagy akár „buDaPEst”-et írunk be, a találatok ugyanazok lesznek.

ÉS (+)

Ezt az operátort (előjelet) alkalmazva arra utasítjuk a keresőt, hogy a találatok között csak olyan hivatkozások szerepeljenek, amelyekben az összes ilyen előjellel rendelkező szó megtalálható.

A keresők többsége alapértelmezetten a megadott keresőszavak mindegyikére keres, nincs szükség az ÉS (+) logikai operátor használatára. Tehát, ha nem használunk semmilyen előjelet, akkor a kereső az ÉS (+) logikai operátort alkalmazza. – Érdeemes megjegyezni, hogy a keresett szavak sorrendje befolyásolja a találati lista hivatkozásainak sorrendjét.

Kivételt képeznek a keresőbe beírt gyakori szavak. Ezeket akkor is kihagyja a keresés szavai közül, ha beírjuk, mert jelentősen lassítanák a keresést, mindazonáltal a keresés szempontjából nem lényegesek. Ilyen például a *hogy*, a *mert* szócskák vagy a névelők. Ezeket csak akkor foglalja a keresendő szavak közé, ha a + logikai operátorral utasítjuk erre a keresőt.

Pl.: [+a +kutya +tartása]

Ne feledjünk szóközt tenni a logikai operátorok elé!

Pl.: [+delfin +élőhely]

Pontos kifejezés (” ”)

Ezt az operátort (előjelet) alkalmazva arra utasítjuk a keresőt, hogy a találatok között csak olyan hivatkozások szerepeljenek, amelyekben a megadott pontos kifejezés minden szava a megadott sorrendben és formában szerepel.

Az ÉS operátor helyett alkalmazhatjuk. Használata egyszerű, csak idézőjelek közé kell tenni a keresett szavak csoportját.

Leginkább akkor érdemes használni, ha pl. személyek nevére ”George Washington” vagy irodalmi szövegekre ”Ott, hol a kis Túr siet beléje” akarunk találatokat kapni.

NEM (-)

Ezt az operátort (előjelet) alkalmazva arra utasítjuk a keresőt, hogy a találatok között csak olyan hivatkozások szerepeljenek, amelyekben a megadott szó nem található meg.

A NEM (-) logikai operátort akkor érdemes használni, ha az általunk megadott keresőkifejezés több értelmű szó, és szeretnénk elkerülni azokat a találatokat, melyek a másodlagos vagy többjelentésből adódnának. Pl.: Ha a Jupiter szóra keresünk, és mi a római istennel kapcsolatos oldalakra akarunk rálelni, akkor: [Jupiter –bolygó]

Ne feledjünk szöközt tenni a logikai operátorok elé!

A kategorizált (tematikus) keresőoldalak, portálok

A kategorizált, vagy tematikus keresőoldalak (angolul *Internet subject directory*) az internetes honlapokat tartalmuk tárgya szerint kategóriákba, csoportokba rendezve kínálja nekünk. A főcsoportokon belül alcsoportokat, alkategóriákat találunk, s ezeken belül listázva láthatjuk a hivatkozásokat. A nagyobb tematikus keresőoldalakon az alkategóriákon belül tovább kereshetünk szabadszavas kereséssel (a szabadszavas keresőoldalakhoz hasonlóan), mivel a hivatkozások listája olyan hosszú, hogy nagyon sok időt igényelne áttekinteni.

Az effajta keresőoldalak előnye az, hogy használóik keresésüket úgy vezethetik végig, hogy tartalmi meghatározások szerint szűkítve keresésüket sokkal rendezettebb és relevánsabb találatokat kapnak. Így az interneten kereső felhasználó csakis olyan találatokhoz jut, mely az kategóriák lépcsős szűkítése során mindegyik tartalmi meghatározás feltételének megfelel.

Például, ha a HuDir (<http://www.hudir.hu>) tematikus keresőben híroldalakat (*hírportálokat*) akarunk megtalálni, akkor a 19 főcsoportból az „Informatika” nevű kategóriára kattintva elénk tárul az ebben a kategóriában található alkategóriák felsorolása. Ez az alcsoport újabb 20 elemből áll, ezek közül kiválasztva a „Számítástechnika” nevű alkategóriát újabb alkategóriákhoz érkezünk, mely további alcsoportokat tartalmaz. Ezt folytatjuk tovább (az alábbi ábra szerint) mindaddig, amíg eljutunk a hírcsoportokhoz.

HuDir

Informatika

Számítástechnika

Internet

Internet hírcsoportok

Fontos megjegyeznünk, hogy a tematikus keresők nem rendelkeznek olyan bőséges honlap-hivatkozás (*link*) adatbázissal, ami megközelítené a szabadszavas keresőkét. Mivel a tematikus keresők kategóriáit és elemeit hozzáértő szakemberek válogatják, nem pedig keresőmotorok, adatbázisuk jóval kisebb, viszont rendszerezett és a honlapok tartalmának fontossága alapján válogatott.