

## A web keresőrendszerei

Az informatikában különbséget illik tenni az angolul *search engine*-nek nevezett keresőrendszerek funkcionális magja, a tulajdonképpeni *keresőgép* és a *felhasználói felület* között. Az átlagfelhasználó szemében viszont nem határolódik el az interfész, a kereső és megjelenítő mechanizmus, valamint az adatbázis. Számára a végeredmény a fontos; a probléma az, hogy hiába a kínálatban szereplő számos kifinomult keresési technika, sokan még a célravezető kulcsszavakat sem találják meg, a kapott választömegből pedig nem ismerik fel, mi a hasznos.

A digitális könyvtáraktól eltérően a keresőszolgáltatásokat főleg a hirdetők finanszírozzák. E szolgáltatások igyekeznek márkahűsége szokatni a keresőket, ezért túlhangsúlyozzák a külsőségeket. Valóban, a keresők általában a használat egyszerűsége, illetve a sebesség alapján kiválasztanak maguknak egy vagy két keresőgépet, és a továbbiakban ehhez ragaszkodnak. Hűségük biztosítására a fejlesztők hallatlan költséggel egyre több soha nem használt eszközzel bővítik rendszerüket.

### A keresőrendszerek története

Az internet a 80-as évek végére vált széleskörűen hozzáférhetővé. A telnet-től a webig vezető úton egyre újabb módszerek kellettek az elérhető információk áttekintésére. Kezdetben szóban, elektronikus levelezéssel vagy sokszorosított listák útján terjedt az új adatbázisok híre, de így nem tudták követni a gyors előrehaladást. Szerencsére egy-egy új internetfunkció megjelenését gyorsan követte valamilyen elektronikus keresőeszköz (FTP:archie; telnet: online tartalomjegyzékek; gopher: veronica; WAIS: keresőszavas tartalomjegyzékek).

A világhálón először 1991-ben lehetett keresni a CERN parancssoros eszközével. 1993 óta vannak Windows- és Macintosh-alapú grafikus keresők, ez visszahatott a WWW áradatszerű növekedésére. Kezdetben elég volt a CERN webhelyén található, ábécérendű tárgymutatóból elindulva végigböngészni a virtuális világkönyvtárat. Amint szaporodtak a HTTP-források, kezdtek megjelenni az eleinte lelkes egyének által fejlesztett, mai keresőgépek. Az évtized közepétől a fejlesztés már iparszerű, és a termékek differenciálódnak.

### A jelen

#### A keresők típusai

Két alaptípus van: tárgyszótárak (Yahoo) és lekerdező robotok (AltaVista, HotBot, Excite). Mindkettő weblapokra mutató adatbázist tart karban. A

tárgyszótárak valamilyen – gyakran meglehetősen bonyolult hierarchiájú – osztályozási rendszer szerint csoportosítják a forrásokat. Az automaták a kereső személy által beírt szöveges kifejezésre épülő algoritmust futtatnak. Mindkét típus alkalmaz a másik logikájából átvett kiegészítő módszereket.

A hálón keresgélő informatikusok próbálnak megbirkózni a túlbujáztott általános és speciális keresőrendszerekkel: osztályozott keresőmotor-jegyzékeket és a keresőkérdést a többi konkrét keresőhöz továbbküldő, a bejövő eredményeket összefésülő metagépeket bocsátanak közre.

### A keresőrendszer tartalma

A weboldal kétféleképpen kerülhet fel a keresőrendszer indexébe: vagy a gazdája jelenti be, pl. az erre kibocsátott weblapon, vagy a keresőrendszernek a linkeket végigkövető, illetve az új és változott lapokat felkutató ügynökszoftverje, robotja deríti és jegyzi fel URL-jét. Az ügynökszoftverek kifinomult stratégiával súlyoznak a bejárható útvonalak között. Az AltaVista pl. napi 6 millió lapot gyűjt be.

A tárgyszótáras rendszerek a felderítést igyekeznek a lényeges forrásokra korlátozni, és a felvétel előtt szűrik a begyűjtött információt.

Az indexelés szolgálatról szolgálatra más elemekre terjed ki az URL-től a teljes szövegig. Az ismertetést olykor kézi úton kulcsszavakkal, osztályozással stb. egészítik ki, ami a minőséget javítja, de a forgást lassítja.

### Keresési lehetőségek

A tipikus nyitó képernyőn van egy input box; a beírt szavakon felül az egyén bejelölhet bizonyos szűkítéseket. Sok keresőgép kínál ingyencenekre való keresőparancsokat, de az eseti felhasználó jobb ha óvakodik ezektől, pl.:

- Boole-algebrai kifejezések (esetenként zárójellezéssel);
- azoknak a szavaknak a megadása, amelyeknek benne kell lenniük, vagy amelyeknek nem szabad benne lenniük a keresett dokumentumban;
- csonkolás;
- teljes kifejezésre keresés;
- helyzeti operátorok;
- korlátozás adatmezőre;
- kis- és nagybetűk megkülönböztetése;
- szűkítés: dátum, nyelv, terület, fájltypus szerint.

### A keresés eredménye

A kapott találatok sokféle formális és tartalmi ismérv (pl. a keresőszavak előfordulásának gyako-

risága, vagy az előfordulás helye, vagy a keresőkérdésbeli súlyozás) szerint rendezve jeleníthetők meg. A keresőgépek egyszerre kb. 10 találatot mutatnak be, és kijelzik a teljes találati halmaz méretét, amiben azután navigálni lehet. A megjelenés formátumát is szabályozni engedik, sőt megkereshetik a találatokban leggyakoribb szavakat, hogy a következő iteratív lépésben ezzel finomítsuk a keresést.

### **Teljesítményértékelés**

Bőségesen léteznek összehasonlító értékelések, de gyorsan elavulnak. A web gyakran aktualizált sajátosság-egybevető táblázatai már jobbák, ám nincsenek elég jól specifikált mérési módszerek. Professzionális mérések is történtek. A vizsgálatokban több rendszer szerepelt: AltaVista, Excite, HotBot, Infoseek, Lycos, Magellan, Open Text, Point. Ezeknek egyszerre tették fel ugyanazokat a kérdéseket, és vizsgálták a keresés sebességét, a találatok mennyiségét, relevanciáját. Az eredmények csupán tájékoztatóak, hiszen eleve nem lehetett nagy a kísérletek száma. Végül is a sztártermékek között nem volt lényeges eltérés, és valahol mindnek volt speciális előnye is, ráadásul ezek gyorsan fejlődnek, hamar elavulttá téve minden összehasonlítást.

### **A jövő**

#### **Indexelés és visszakeresés**

A téma kutatói rámutatnak, hogy a keresések gyengése inkább az indexelendő dokumentumok indexelhetetlenségén, mint a keresőrendszere tehetségén múlik. Azt javasolják, hogy a mostani technikák finomítása helyett – ezek ugyanis homogén és jól strukturált dokumentumvilágról álmodnak – vegyék tudomásul, hogy az átlagkereső egykét keresőszóval, semmi mással nem akar boldogulni. Néhány javasolt fejlesztési irány:

- intelligens ügynökszoftverek,
- természetes nyelvi indexelési módszerek,
- szabályalapú forrásfelderítés,
- képi és hangdokumentumok automatikus indexelése.

#### **Felhasználói beállítások**

Ha ismerik a felhasználók csoportjait, igényeiknek megfelelőbb kinézetű, szakterületükön használt paramétereket bekérő, csak a releváns adatbázisokban és csak az ő logikájuk szerint kereső, ezáltal hatékonyabb keresőrendszereket lehet kifejleszteni. Az általános célú keresők fel is ajánlanak ilyen felhasználói opciókat. Némely keresőrendszer felajánlja, hogy ha a találatok közül a felhasználó kiemelte a relevánsakat, akkor hason-

lóság szerint automatikusan tovább keres, egyre jobb találatokért.

### **Összesítés**

A keresőrendszerektől elvárható, hogy ne csak rendezve, de kategorizálva adják vissza a találatokat.

### **A keresés kiterjesztése**

Néhány keresőrendszer rendelkezésre bocsátja a találatok dokumentumaiban leggyakrabban előforduló szavak rangsorát, hogy a következő iteratív keresésben a felhasználó ezek közül egészíthesse ki kérdését.

### **Hatókör**

A világhálón szaporodó multimédia adatformátumok nem hagyhatók figyelmen kívül. Ezek felismerése még nem tökéletes, az indexelésükhöz viszont már vannak jó ujjlenyomatkepző technikák.

### **Metaadatok**

Ezek a bibliográfiai leírásra emlékeztető, helyettesítő dokumentumok az elsődleges adatbázisok felépülése során automatikusan előállhatnak, ha a dokumentumok kategorizálhatók, és a kategórián belül fegyelmezett formátumuk van. A keresés ilyenkor a metaadatokat tekinti végig. E módszer nemigen vihető át az általános keresőszolgáltatás világába, de tudományos, vagy más behatárolt felhasználói körben nagyon hatékony. Az általános keresők elé kapcsolt metakereső (Infoseek) párhuzamosan több keresőrendszer számára is konvertálja a kapott kérdést, az eredményeket pedig normalizálás után, összefésülve adja vissza.

### **Következtetések**

Az információ-visszakeresés történelmi fordulópontján állunk. A mai számítógépekkel elfogadható időn belül óriási adatbázisokból meglehetősen bonyolult algoritmusokkal elég teljes és jó minőségű találati listát kaphatunk. A nagy adatbázisokat birtokló gazdálkodó szervek érdekeltek az információpiacon, támogatják a keresőrendszerek fejlesztését. Költségvetési pénzekből fejlődhet a digitális könyvtárak világa. A tudományos és üzleti szféra céljai közel kerültek egymáshoz.

/SCHWARTZ, Candy: Web search engines. = Journal of the American Society for Information Science, 49. köt. 11. sz. 1998. p. 973–982./

(Góth László)

## **6000 ELEKTRONIKUS FOLYÓIRAT AZ OMIKK-BAN!** (SZAKFOLYÓIRATOK TELJES ANYAGA A SZÁMÍTÓGÉPBŐL)

Könyvtárunkban térítésmentesen használhatja az alábbi szolgáltatásokat:

### **Electronic Information for Libraries (EIFL)**

*az EBSCO Publishing rendszere a Soros Alapítvány Nyílt Társadalom Intézete programjában*  
**3000 szakfolyóirat és 143 hírlap** teljes számai és archívuma négy adatbázisból (Academic Search Elite, Business Source Premier, MasterFILE Premier, Newspaper Source).

### **Elsevier – ScienceDirect**

*az Elsevier Science kiadói rendszere*

Valamennyi - több mint **1000** - Elsevier-folyóirat teljes számai és archívuma (az OMIKK által előfizetett 65 folyóiraton kívül a letöltések száma korlátozott).

### **SpringerLink**

*a Springer kiadói rendszere*

Több mint **400 Springer-folyóirat** elektronikus változata. A Springer-lapok a **SewtsNeten** is hozzáférhetők, ahol **15 000 folyóirat** tartalomjegyzékei, cikkösszefoglalói érhetők el.

### **ProQuest Applied Science and Technology Plus**

*a Bell and Howell Information and Learning webes szolgáltatása*

**648 folyóirat** bibliográfiai adatai és cikkösszefoglalói, ebből **297 folyóirat** teljes számai és archívuma, hatékony, felhasználóbarát keresőrendszerrel.

### **Business and Industry**

*a Responsive Database Services, Inc. OnDisc adatbázisa*

A neves, klasszikus üzleti adatbázisban több mint **900 folyóirat** teljes szövegű cikkei.

### **EBSCO Online**

*az EBSCO Information Services online rendszere*

**4000 folyóirat** tartalomjegyzékei, cikkösszefoglalói, köztük **16 folyóirat** teljes elérése

### **Computer Select**

*a GML Corporation adatbázisa*

**69 számítástechnikai folyóirat** teljes szövege.

### **Elektronikus folyóiratok közvetlen elérése kiadói webről**

Az OMIKK termináljairól jelenleg **12 folyóirat** elektronikus változata érhető el közvetlenül a kiadók weblapjairól.

**További információért kérjük, forduljon a Tájékoztató szolgálat munkatársaihoz, személyesen, telefonon (338-4089) vagy e-mailen (refposta@omk.omikk.hu) keresztül.**