

ONLINE INFORMÁCIÓKERESÉS

Online keresés az ESA–IRS rendszer
adatbázisaiban szabályozott és
szabad tárgyszavakkal

*Az Európai Űrkutatási Szervezet online
információkereső rendszere*

A 70-es évek elején alakult meg az *Európai Űrkutatási Szervezet (European Space Agency, ESA) Információkereső Szolgáltatása (Information Retrieval Service, IRS)*. Az ESA–IRS RECON elnevezésű információkereső rendszere interaktív módszerrel, online kapcsolat útján teszi lehetővé a Frascatiban (Róma mellett) működő központi nagy számítógépben tárolt adatbázisok használatát, közvetlen lekérdezését. Az adatbázisokhoz az ESANET hálózaton keresztül lehet hozzáférni úgy, hogy a felhasználói terminálok és a központ (vagy valamely csomópont) között telefonvonalakon (tárcsázással) létesül kapcsolat a keresés idejére.

A RECON keresőrendszer legfontosabb „parancsai” a következők:

- BEGIN:** a keresésre kiválasztott adatbázis előhívása;
- EXPAND:** az adatbázis „szótárából” a kívánt alfabetikus részlet kiírása a képernyőre;
- SELECT:** a szótárból kiválasztott keresőszavak aktivizálása és adatbázisbeli gyakoriságának kérése;
- COMBINE:** a kiválasztott keresőszavak kombinálása Boole-algebrai operátorokkal (ÉS, VAGY, NEM);
- DISPLAY:** az előző parancsnak elegendő információ tételek megjelenítése a képernyőn.

A fenti parancsokon kívül – persze – mások is a rendelkezésre állnak. Ez utóbbiak – a fő parancsokat módosító mellékutasításokkal együtt – a keresés finomítását, pontosítását és egyéb körülményeinek behatárolását teszik lehetővé.

A keresési stratégia menete négy lépésre bontható:
a keresőszavak kiválogatása;
ezek kombinálása;
az eredményként kapott „találatok” kiírása és értékelése,
az első stratégia módosítása.

Az első lépésben a keresést végző személy az EXPAND és a SELECT paranccsal kiválasztja a témára jellemző kulcsszavakat, akár az adatbázis tezaurusának

deszkriptorai („szabályozott szókincs”), akár a címben vagy a referátumban használt szakkifejezések (szabad tárgyszavak), akár a bibliográfiai vagy egyéb adatok közül. A második lépésben ezeket a COMBINE paranccsal kombinálja, vagyis meghatározza egymáshoz való viszonyukat. A keresőt mindkét lépésben a gép tájékoztatja arról, hogy a bevitt keresőszavakat ill. kombinációikat hány információs tétel tartalmazza. A harmadik lépés a keresés kritériumait kielégítő, vagyis relevánsnak ígérkező tételek egymás utáni kiírása a terminál képernyőjére. Ez lehetővé teszi további szabályozott vagy szabad tárgyszavak beépítését a keresésbe, egyesek kizárását vagy a kombináció megváltoztatását. Tehát a relevancia javítása céljából a második lépésben definiált kezdő keresési stratégiát módosíthatjuk, az itt vázolt interaktív vagy „párbeszédes” módszerrel.

Az indexelés

Indexelésnek nevezzük azt az intellektuális műveletet, amellyel egy-egy dokumentumhoz olyan tárgyszavakat rendelünk hozzá, amelyek egy adott dokumentum tartalmát tükrözik. Ezek többnyire valamely tezausból vagy szabályozott tárgyszójegyzékből kiválasztott egyszerű szavak vagy összetett szakkifejezések – egységes nyelvtani alakra hozva. Némely adatbázis – a szabályozottakon kívül – megengedi a szabad tárgyszavak hozzárendelését is. Ezeket a legtöbb esetben a dokumentum szövegéből veszik át.

Az adatbázis előállításának másik intellektuális feladata a referátumok megírása. A jó referátum és a tartalmat híven kifejező (esetleg kiegészített) cím hozzájárul a releváns információk hatékony kikereséséhez.

Az ESA–IRS adatbázis-központban mintegy 20 bibliográfiai adatbázishoz lehet online módon hozzáférni, ami a műszaki és természettudományok széles körű lefedését teszi lehetővé. Az adatbázisok indexelési szintje és módszere igen változatos. A rendszerben levő 14 legfontosabb adatbázist és hozzáféréseinek tartalmi eszközeit az *1. táblázat* foglalja össze. Ebből látható, hogy két adatbázisnak nincs szabályozott szókincsre alapozott indexelése: a CHEMABS (Chemical Abstracts) informáci-

ós rekordjaihoz számos szabad indexkifejezést rendelnek hozzá; a BIOSIS-nak is szabad tárgyszavai vannak, de ezeket bizonyos nyelvtani megkötésekkel szabad csak használni.

A többi adatbázisnak szabályozott indexelési rendszere, ill. szókincse van, közülük azonban hat szabad tárgyszavazást is lehetővé tesz.

A további különbségek az indexelés mélységéből adódnak, pl. a NASA adatbázis rekordjaihoz átlagosan 10 deskriptor tartozik; az INSPEC rekordokhoz némely esetben egy, máskor öt vagy még több deskriptor, továbbá átlagban 7 szabad tárgyszó vagy szakkifejezés is; más adatbázisok tételeinek átlagosan 8 szabályozott deskriptora, de csak egy-két szabad tárgyszava van, és így tovább.

1. táblázat

Az ESA-IRS rendszer fontosabb
adatbázisainak kereshető adatmezői

	Szabályozott tárgyszavak	Szabad tárgyszavak	Cím	Referátum
NASA	+		+	
CHEMABS		+	+	
METADEX	+		+	
COMPENDEX	+	+	+	
NTIS	+	+	+	
BIOSIS		+	+	
INSPEC	+	+	+	
ALUMINIUM	+		+	+
ISMEC	+	+	+	
ENVIROLINE	+		+	+
PASCAL	+		+	
OCEANIC	+	+	+	+
POLLUTION	+	+	+	+
ENERGYLINE	+		+	+

Az invertált fájl

Az ESA-IRS központba beérkező valamennyi, gépi adathordozóra rögzített adatbázisból (pontosabban azok növekményeiből) azonnal invertált fájlokat készítenek. Ez azt jelenti, hogy az egyes tételek minden adatmezőjének (a szerzők, a cím, a referátum, az indexek stb.) valamennyi adatából (pl. szavaiból) „szótárakat” állítanak elő. A szótár minden címszava mellett megtalálható, hogy hányszor fordul elő az adatbázisban; a gép – a SELECT parancsra – ki tudja keresi azokat a tételeket, amelyekben a kijelölt szavak előfordulnak. Az invertált fájl címszavai alfabetikus rendezettek. Az indexkifejezések minden szava külön-külön is előfordul a szótárban,

de összetett szakkifejezések formájában is szerepel benne. Az EXPAND paranccsal éppen az invertált fájl szótárának tetszőleges részletét tudjuk előhívni a képernyőre.

Mivel a keresés az indexelés fordított műveletének is tekinthető, ezért meg kell ismerkednünk a dokumentum tartalmának kifejezésére szolgáló főbb módszerek elveivel.

Indexelés deskriptorokkal

Az indexelő nyelvek elsődleges célja az információkeresés teljességi ill. pontossági mutatójának javítása. A teljesség növelésére szolgáló eszközök *növelik* a kikeresett dokumentumok számát, szélesítik a keresőszó értelmezését, csökkentik az indexelő és a kereső-szókincs nagyságát. Ilyen eszközök: a szinonima-szabályozások, a hierarchikus kapcsolódások, a szóalakok egységesítése stb. A pontosság növelésére szolgáló eszköz pl. a koordinálás, a kapcsolódó kifejezések megjelölése, a szerepjelölés, a súlyozás; ezek viszont *csökkentik* a dokumentumhalmaz nagyságát és növelik a kereső-szókincset.

A legtöbb teljesség- és pontosság-fokozó eszközt beépítették a teauruszokba vagy a szabályozott tárgyszórendszerekbe. A NASA adatbázis teauruszában pl. a szinonimákat is szabályozzák és a pontosság növelésére pre-koordinációt is alkalmaznak (pl. a PSYCHOACOUSTICS – pszicho-akusztika vagy az AGROCLIMATOLOGY – agrár-klimatológia deskriptorokkal). A teauruszba sokféle olyan kapcsolatot is beépítettek, amelyek sem szinonima-, sem hierarchikus kapcsolatot nem fejeznek ki (pl. antonimia, anyagra, eszközre utaló, hasznosítási kapcsolatok). Ezeket a kapcsolatokat is – a szinonimákhoz hasonlóan – a USE és a USED FOR (UF) jelöli. Például:

antonimia:

INSTABILITY USE STABILITY,

instrumentális:

SUPPLIERS USE SUPPLYING

(a fordított kapcsolatokat az UF mutatja).

A jól kidolgozott teauruszok ellenére az indexelés sok szubjektivitást hordoz magában. Az indexelési következetességet olyan tényezők is befolyásolják, mint a teauruszok korlátai, a nyelvi és szintaktikai többletműség, a szerző-indexelő „interface” (a szerző gondolatának átvétele, szellemi rezonanciája), az indexelés vezérelve, az indexelő tárgy- és nyelvismerete, az indexelő-kereső „interface”.

A legfontosabb indexelési-információkeresési szabály: az indexelőnek mindenkor arra kell gondolnia, hogy a kereső hogyan találhatja meg az indexelt dokumentumot, és megfordítva, a keresőnek azt kell kitalálnia, hogyan indexelték az általa keresett dokumentumokat.

Indexelés szabad tárgyszavakkal

A teaurusz deskriptoraitól eltérő tárgyszavakkal végzett „szabad indexelés” a visszakeresés határfokának javítását szolgálja. Egyes esetekben, mint pl. az INSPEC adatbázisban, a szabad tárgyszavak kiegészítik a szabályozott deskriptorokat, és a szerzők szóhasználatát tükrözik (az indexelő a közlemény címéből és referátumából kölcsönözte őket).

A szabad tárgyszavak használata semmiféle előírást nem követ. A kereséskor növelheti a teljességet, mert a kereső személy részben a saját szóhasználatára is építhet, részben olyan keresőszavakat is használhat, amelyek a teauruszból esetleg hiányoznak. A teauruszok általában bizonyos időkésséssel követik az új műszaki és tudományos fogalmak bevezetését, ezért ilyenkor szinte kizárólag a szabad tárgyszavakra kell építenünk a keresésben.

Az információkeresés szempontjai

Mielőtt részletesen összehasonlítanánk a szabályozott szókincsre és a szabad tárgyszavakra alapozott keresés előnyeit és hátrányait, vizsgáljuk meg röviden azt, hogyan teszik alkalmassá az invertált fájl szótárait az online keresésre.

A szabályozott és a szabad tárgyszavak, a cím és a referátum szavai alfabetikus rendben hívhatók le a képernyőre az EXPAND utasítással. Az ESA-IRS rendszerben az INSPEC és a NASA adatbázisok esetében a szabályozott deskriptorok hierarchikus alá- és fölérendeltségi kapcsolatait, valamint a kapcsolódó kifejezések (related terms) is előállíthatók, sőt az INSPEC esetében még a USED FOR relációk is. A többi adatbázisban, ha van is teauruszuk, a hierarchikus kapcsolatok nem mutathatók ki online kereséskor, viszont a NASA teauruszának hierarchiája és kapcsolódó kifejezései segítőeszközként esetükben is használhatók.

A deskriptorok és az összetett szabad tárgyszavak (szakkifejezések) kétféleképpen szerepelnek a szótárban: unitermként és multitermként. Pl. a SATELLITE ATTITUDE CONTROL deskriptor így, összefüggően is megjelenik az S betűnél (multiterm), de megjelenik három diszkrét unitermként is: SATELLITE, ATTITUDE, CONTROL, mindegyik a maga helyén. A *Chemical Abstracts* adatbázis kivétel, a szabad tárgyszavakat a szótár csak unitermként tartalmazza. Megjegyzendő, ha valamely tárgyszó az adatbázisban csak egyszer fordul elő, nem veszik fel a szótárba, nehogy túl nagyra nőjön az invertált fájl.

A tárgyszavak kettős szerepeltetésének (unitermek és multitermek) jól átgondolt okai vannak; mindkét fajtnak meg kell lennie ahhoz, hogy a visszakeresés maximális határfokú legyen. A cím és a referátum szövegének

jelentős szavai (a kötőszók, előljárószók, névelők stb. elhagyásával) csak unitermekként szerepelnek a szótárban.

*Keresés szabályozott tárgyszavak (deskriptorok) alapján**Előnyök:*

a kereső további deskriptorokhoz juthat el a teaurusz hierarchikus szerkezete és kapcsolódó kifejezései révén;

nem kell a szinonimákat is kitalálni;

kevesebb online kapcsolt idő szükséges, ha nyomtatott teauruszt használ a kereső;

a szabályozott szókincs segíti a tárgykörben nem kellően jártas keresőt azzal, hogy nem kell utánajárnia a valószínű keresőszavaknak;

nagyobb keresési pontosság, kisebb zaj érhető el;

a keresőnek nagyobb esélye van az indexelő által használt precíz tárgyszavak megtalálására;

a visszakeresés akkor is lehetséges, ha történetesen a dokumentumnak figyelemfelkeltő, nem szakszerű címe van.

Hátrányok:

sok időbe telik a teaurusz kellő ismeretének megszerzése;

a deskriptorok használata az indexelő szubjektív ítéletétől függ;

a szabályozott szókincs használatát az indexelés vezérelvei, előírásai megkötik, és ez csak akkor változik, ha új deskriptorokkal egészül ki;

a keresőkérdést a közvetítőnek indexnyelvre kell lefordítania, ami eltérhet attól, ahogyan ő értelmezi a kérdést;

minden adatbázisnak külön-külön saját szabályozott szókinccse van.

*Keresés szabad tárgyszavak alapján**Előnyök:*

a kereső azonnal használhatja az általa ismert szakkifejezéseket;

növekedhet a teljesség, ha a deskriptorok nem tükrözik pontosan a dokumentum tartalmát;

azonos keresőszavak, keresőkifejezések használhatók minden adatbázisban.

Hátrányok:

nem ismeretesek a keresőszavak közötti összefüggések – igaz viszont, hogy az alfabetikus szótárból ellenőrizni lehet az egyes szavak írásmódját;

ritkaság, hogy egyetlen egyértelmű kifejezéssel végezhető el a keresés – több szinonimát vagy rokonértelmű kifejezést kell a keresés közben kigondolni;

nagyobb a nem releváns találatok esélye, mert a pre-koordinálatlan uniterm keresőszavakat különböző szöveggörnyezetből vesszük ki;

nagyobb a veszteség esélye, ha a cím nem szakszerű; veszteség adódhat abból is, ha a szabad szöveg multiterm szakkifejezéseit nem felbontva keressük;

a szabad tárgyszavak többértelműségével, következetlenségeivel, rövidítéseivel kell számolni.

Az információkeresés sikere erősen függ a különféle megközelítési módszerek alkalmazásától és a használt szókincs ismeretétől. Ha nem áll rendelkezésre szabályozott szókincs, a keresőnek nagyon megnehezülhet a dolga. Más kérdés, hogy ilyenkor egy másik adatbázis teauruszát hívhatja – kiindulásként – segítségül.

Vannak olyan információs tételek, amelyek szabad tárgyszavakkal (pl. csak a cím szavaira hagyatkozva) sohasem kereshetők vissza; ugyanezek szabályozott deskriptorok alkalmazásával könnyen kikereshetők (pl. egy másik adatbázisból). Más esetekben viszont a deskriptorok használata elégtelen; növeli a keresés teljességét, ha a deskriptorokon kívül pl. a címbeli tárgyszavakkal is próbálkozunk.

A szabad tárgyszavas keresés további hátránya, ha a címben rövidítéseket (pl. vegyjeleket) használnak. Legyen egy ilyen cím: *Cr és Si meghatározása acélban*. A kereső hiába viszi be a króm és a szilícium keresőszavakat, a dokumentumot így nem találja meg; deskriptorok esetében igen.

Egy szabályozott tárgyszavas rendszerben a *repülőgép* fogalmat csak egyféleképpen kell keresni: AIRCRAFT. Szabad tárgyszavas kereséshez viszont a következő alternatívákra kell gondolni: AIRCRAFT, AIRPLANE, AIRPLANES, AEROPLANE, AEROPLANES, PLANE, A/C. Másrészt azonban egyazon alakú szónak különböző jelentései lehetnek; pl. a PLANE jelentései: repülőgép, gyalu, platánfa, színvonal, síklap, mértani sík. A szinonimák és homonimák szövevénye teljesen meghamisíthatja egy-egy keresés eredményét, akár a teljesség, akár a pontosság vonatkozásában. Különösen igaz ez akkor, ha csak a címbeli szavakra végezhetünk keresést, mint pl. az ISI (Institute for Scientific Information) adatbázisa esetében, amelyet éppen ezért kivettek az ESA–IRS online módon hozzáférhető adatbázisai közül.

Keresési stratégiák

Az alapvető online keresési stratégiák nemigen térnek el egymástól, akár szabályozott, akár szabad szókincsű keresést végzünk. Mindkét esetben lehívható a tárgyszavak alfabetikus jegyzéke, sőt egyes esetekben a hierarchikus szerkezet is. Ennek hiányában a keresőnek kell kigondolnia a generikus és a kapcsolódó kifejezéseket, hacsak nincs teaurusza.

Szabályozott terminológia esetében nem szükséges a tárgyszójegyzéket lehívni, de ebben az esetben ki vagyunk téve annak, hogy néhány újabb deskriptorról nem szerzünk tudomást (különösen akkor, ha régi kiadású teaurusz áll a rendelkezésünkre). Szabad tárgyszavas keresés esetében azonban nagyon tanácsos az EXPAND parancs használata, hogy a szavak különféle alakjairól, írásmódjáról, helyesírásáról meggyőződhesünk.

A keresőszavak szempontjai szerint az online keresési stratégiák fő válfajai a következők:

váz-kialakítású stratégia, amelyben először csak néhány (szabályozott vagy szabad) tárgyszót viszünk be, mint vázat, és a kapott válaszokból – új szavak bevonásával – tovább építjük a stratégiát;

teljességre törekvő stratégia, amelynek során azonnal bevisszük az eszünkbe jutó összes keresőszót;

gyors, de nagy „zajt” termelő stratégia, amelyben sok szakkifejezést csonkolva viszünk be unitermekként, egymással kombináljuk őket, majd a válaszokat – tartalmukat mérlegelni sem próbálva – rögtön kinyomtatjuk.

Az ilyen stratégiákban kétfajta közelítéssel dolgozhatunk:

1. az egyszerű szavakból álló halmazokból metszetet képezünk (mint a gyors módszerben);
2. kontextuális logikai közelítést alkalmazunk.

Az utóbbi eljárással a kereső kijelölheti azt, hogy a keresőszavak a rekord azonos adatmezőiben (címben, referátumban stb.) vagy a referátum azonos mondatában forduljanak elő (ld. később, a 2. táblázathoz adott magyarázatban is). Ilyenkor a keresőszavak bármilyen sorrendben lehetnek, vagy meghatározható a közöttük megengedett egyéb szavak száma.

A kontextuális logikai módszer előnye, hogy csak egy „keresőmondatot” kell képezni, amely több fogalmat is magában foglal. Ha pl. keresőkérdésünk a következő:

A repülőgépzaj (aircraft noise) káros hatása a nők (women) alvására (sleep),

akkor a kontextuális módszerrel kialakított keresési stratégia egyetlen sor:

AIRCRAFT(W)NOISE(F)SLEEP(F)WOMEN

A W kapcsolójel azt jelenti, hogy a két szomszédos szó között nem lehet másik, az F pedig azt, hogy azonos adatmezőben kell az egy multiterm és a két uniterm tárgyszónak előfordulnia, de tetszőleges sorrendben.

Az 1. alatti stratégia esetében három különböző szócsoporthoz (AIRCRAFT NOISE, SLEEP, WOMEN) kellene képeznünk és ezeket kombinálni ÉS kapcsolattal. Van amikor az 1., s van, amikor a 2. megoldás a kedvezőbb.

A keresés teljességét nem befolyásolhatja az a tény, hogy a kereső szó szabad tárgyszó-e vagy szabályozott deskriptor. A kontextuálisan vagy egyszerű szavak kombinációjaként kialakított tetszőleges kereső-kifejezések több találathoz vezetnek, mint a kizárólagosan a

deszkriptorokra alapozott keresés. Ennek oka, hogy az előbbi esetekben egyaránt keresünk a cím, a referátum és a deszkriptorok adatmezőjében előforduló tárgyszavak szerint, az utóbbi esetben viszont csak a deszkriptorok szerint. Ezt bizonyítja a 2. táblázat, ahol a repülőgép által okozott zaj (AIRCRAFT NOISE) kereső-kifejezést különféleképpen keressük egy adatbázisban.

2. táblázat
Az AIRCRAFT NOISE szakkifejezés keresése
különféle módszerekkel

A kérdés sor-száma	A találatok sor-száma	Kereső-kifejezés	Magyarázat
1	2250	AIRCRAFT NOISE	csak szabályozott deszkriptor
2	3662	AIRCRAFT(W)NOISE	a kifejezés bármely adatmezőben előfordulhat, de csak egymás mellett, ilyen sorrendben
3	4979	AIRCRAFT(F)NOISE	a tárgyszavak ugyan-csak bármely adatmezőben előfordulhatnak, de nem kell egymás mellett állniuk
4	5148	AIRCRAFT(C)NOISE	a tárgyszavaknak nem kell ugyanabban az adatmezőben sem lenniük
5	55 906	AIRCRAFT	logikai kombináció, a két tárgyszó bárhol előfordulhat a rekordban, az eredmény ugyanaz, mint a 4. számú kérdésre
6	30 074	NOISE	
7	5148	5 ÉS 6	

Látható, hogy mindegyik kontextuális logikai módszerrel feltett kérdés (2., 3. és 4. kérdés) tartalmazza az 1. kérdésben hivatkozott deszkriptort is. Valószínű, hogy a leginkább releváns dokumentumokat az 1. kérdés szolgáltatja, mert az indexelők ezt a deszkriptort nyilván azért rendelték hozzá a dokumentumhoz, hogy tükrözze annak fő mondanivalóját. Ilyenkor persze elveszítjük azokat a dokumentumokat, amelyek az AIRCRAFT NOISE szakkifejezést a címben vagy a referátumban tartalmazzák, de ilyen deszkriptoruk nincs. Ezeket (a deszkriptoros tételekkel együtt) a 2. kérdés „hozza ki”. Ha viszont a 3. és 4. kérdés kontextuális közelítését vagy az 5–6–7. kérdések egyszerű szavas, kombinációs közelítését alkalmazzuk, hamis válaszokat is kaphatunk, mert az AIRCRAFT és a NOISE nem feltétlenül egymáshoz rendelt szavak lehetnek.

A keresési stratégiát bizonyos fókig befolyásolja az indexelés. Ha az információ tételeket egyáltalán indexelték, azaz intellektuális munkával valamilyen tárgyszavakkal látták el, akkor a keresőnek az indexelő fejével kell gondoskodnia: ki kell találnia, hogyan indexelhették az általa keresett dokumentumokat. Ha ez nem vezet eredményre, akkor a témát olyan kérdésekkel kell megközelítenie, amelyek segítenek neki a releváns keresőszavak kikutatásában.

A szabályozott és szabad tárgyszavas online keresési módszerek közül a keresőnek kell választania, az előnyök és hátrányok alapos ismeretében. Döntésében olyan tényezők is szerepet játszanak, mint az online keresésre fordítható, rendszerint korlátozott idő, a kívánt válaszok száma, a keresés pontossága vagy teljessége.

/RAITT, D. I.: *Aspects of searching via on-line systems using controlled and uncontrolled vocabularies.* = *IATUL Proceedings, Online issue, 12. köt. 1980. p. 3–21./*

(Roboz Péter)



Egy újabb stratégia az információkeresésben

Bevezetés

A működő információkereső rendszerek (Information Retrieval System, IRS) száma a százat is meghaladja; közös céljuk: ellátni az olvasókat igényeik szerinti dokumentumokkal. Ami megkülönbözteti ezeket az IRS-eket az adatkereső rendszerektől (Data Retrieval System, DRS), amilyenek pl. a bankügyleti vagy légiforgalmi helyfoglaló rendszerek, az a *relevancia*. A DRS-ek esetében egy kérdésre kapott válasz relevanciája *objektív módon dönthető el*, míg az IRS-ekben a feltett kérdésre kapott válasz relevanciáját voltaképpen csak maga a felhasználó tudja megítélni. A DRS-eket – e szempontból – az IRS-ek speciális eseteként tekinthetjük. A relevancia az a tényező, amely meghatározza az IRS-ek hatékonyságát.

A működő IRS-ek, amelyek közül egyesek a nagy, online hozzáféréssű adatbázisok keresését is lehetővé teszik, azonos alapon működnek: a felhasználó kérdését deszkriptorok és Boole-operátorok segítségével keresőprofilá alakítják át, és összehasonlítják a fájlban tárolt dokumentumokhoz hozzárendelt deszkriptorokkal. A fájl minden dokumentuma egyenként átesik ezen a