

Ungváry Rudolf, Vajda Erik

Könyvtári információkeresés

Előszó

Ezt a könyvet elsősorban könyvtárosoknak és – kissé talán elavult kifejezéssel – szakirodalmi tájékoztatással foglalkozó szakembereknek, továbbá leendő információközvetítőknek és brókereknek vagy e hivatáskörre készülőknek írtuk, akiknek tanulmányaikhoz, továbbképzésükhöz és gyakorlati munkájukhoz kívántunk segítséget nyújtani. Mindez nem jelenti azt, hogy a könyv nem tehet (és reméljük tesz) hasznos szolgálatot az információk “végfelhasználóinak”, akik az információs technológia fejlődése révén egyre gyakrabban találkoznak azzal a lehetőséggel és egyben feladattal, hogy maguk keressék a szükségleteiknek megfelelő információkat. Ez különösen vonatkozik a könyvnek az interneten végezhető információkereséssel foglalkozó részeire, de azt hisszük, hogy a könyv tartalmának nagy része hozzásegíti az információhasználókat vagy ahhoz, hogy maguk éljenek az információkeresésre szolgáló források és eszközök használatának lehetőségeivel, vagy ahhoz, hogy – felvilágosult fogyasztóként – tudják, mit várhatnak az információs szolgáltatás “főfoglalkozású” munkatársaitól, illetve mivel járulhatnak hozzá az utóbbiak munkájának eredményességéhez.

A könyv témájául választott szakterületnek mind gyakorlati alkalmazási háttere, mind eszköztára, mind pedig – az előbbiektől nem függetlenül – ismeretanyaga és irodalma az általában is gyors fejlődésnél is gyorsabban fejlődik, növekszik. Nem törekedhettünk arra, hogy e nagy ismeretanyagot teljes körűen feldolgozzuk, és tartalmát teljes részletességgel tárgyaljuk. Tisztában vagyunk azzal, hogy az információkereséssel hivatásszerűen foglalkozó, nagy gyakorlati tapasztalatokkal rendelkező információközvetítőknek ebben a könyvben kevés újat mondhattunk. Mégis, abban bízunk, hogy a fontosabb, könnyen továbbfejleszthető ismereteket ez a könyv közli, amellett, hogy nem nagy, – egy esetben gyűjteményes munkára is utaló – irodalomjegyzéke is segíthet további ismeretek megszerzésében.

Az információkeresés gyakorlati szakterülete és eszköztára igen gyorsan fejlődik, nem szólva arról, hogy már ma is igen sokrétű. Hangsúlyozni kell, hogy a könyv szerzőinek minden jó szándéka és a könyv bizonyára számottevő tartalma ellenére, különösen a professzionális információkeresőknek a gyakorlatban is meg kell tanulniuk, hogy a rendelkezésre álló, illetve felderíthető szolgáltatásokkal milyen esetben érdemes élni, és milyen módszereket kell használni.

Az információkeresés az esetek többségében “társasjáték”, amiben a végfelhasználónak és a közvetítő szakembernek megvan a maga szerepe. Ez mindkét részről bizalmat és türelmet igényel. A professzionális információkeresők (köztük kiemelkedő számban és szereppel a könyvtárosok) nem elégedhetnek meg azzal, hogy a végfelhasználók elé tárják az információkeresésre szolgáló eszköztárat. Gondoskodniuk kell arról is, hogy a végfelhasználónak – ha nincs lehetőség vagy szükség arra, hogy helyette elvégezzék az információkeresés munkáját – segítséget nyújtsanak ennek az eszköztárnak a használatában. A könyvtári tájékoztató szakembernek (közkeletű nevén referenz-könyvtárosnak) is, az információkeresésre specializálódott – általában szakterületi – munkatársaknak is néha kéretlenül is hozzá kell segítenie a végfelhasználót ahhoz, hogy megtalálja, amire szüksége van, még ha nincsenek is teljesen azonos feladataik és lehetőségeik az információkeresésre szakosodott és arra vállalkozó információbrókerekkel. Reméljük, hogy ez a könyv nem csak meggyőzi erről az információkeresésre, tájékoztatásra szakosodott szakembereket, hanem alkalmas lesz arra is, hogy a végfelhasználók – növekvő ismereteik birtokában is – igénybe vegyék a közvetítői segítséget.

Utoljára, de nem utolsó sorban megköszönjük Murányi Péternek és Vadász Ágnesnek észrevételeiket és javaslataikat, melyeket e könyv második, javított kiadásához tettek.

Budapest, 2002. január

A szerzők

1. Könyvtár és információkeresés. Áttekintés

1.1 A könyvtári funkció és jövője

Maga a könyvtári funkció, vagyis a könyvtár léte, hosszú, több évezredes története, széleskörű használata, számos esetben nélkülözhetetlensége és társadalmi elismertsége az alábbiakból következik:

- a) A gondolatokat, ismereteket és érzéseket hangokkal, mozdulatokkal, gesztusokkal csak akkor adathatjuk át és foghatjuk föl, ha a közlő és a fogadó időben és térben egyaránt olyan közel vannak egymáshoz, hogy e hangokat, mozdulatokat, gesztusokat érzékelhetik. A kommunikáció említett formáin alapuló közvetlen információátadás az időbeli és térbeli közelség nélkül megvalósíthatatlan.
- b) Maga a közlő sem rendelkezik az időben korlátlanul gondolataival, ismereteivel és érzéseivel, részben az emberi lét és az emberi emlékezet korlátozottsága folytán.
- c) Az ember mint leszármazottsági, pár- és csoportkapcsolatokban és mint munkamegosztásra alapozott társadalomban élő lény szükségképpen közölni akarja (közölni kénytelen) gondolatait, ismereteit, érzéseit. Az emberi lét fel is tételezi ezek közlését, másfelől pedig az egyes ember és a társadalom rászorul arra, hogy e közléseket fogadni és tárolni tudja.
- d) Azok, akiknek kívánsága, érdeke, sőt létérdeke a közlés és mások közléseinek fogadása, térben és időben gyakran olyan távol vannak egymástól, hogy az eredményes közlés, illetve a közlés eredményes fogadása lehetetlen. A távközlés technikája ugyan lehetővé teszi a közlést és annak fogadását a térben távoliak között, de a közölt információk rögzítése nélkül ez csak egyidejűleg mehet végbe.
- e) Az időben és – többnyire – térben is távol esők között az információcsere megköveteli az információk rögzítését, vagyis dokumentumok létrehozását.
- f) A dokumentumok – csakúgy, mint a nem rögzített információk – a térbeli és időbeli távolság okán nem szükségképpen jutnak el mindenkihez, aki az adott információ fogadásában érdekelt.
- g) Még ha a térbeli és időbeli távolság nem is jelentene korlátot, a dokumentumok használata azért is nehézségekbe ütközik, mert feltétele, hogy a fogadó (nevezzük olvasónak) tudjon a dokumentum létéről és hollétéről, meg tudja állapítani, hogy szüksége van rá, és hozzá is férjen.

És ezzel el is jutottunk a könyvtári funkció lényegéhez. **A könyvtári funkció abban áll, hogy az célra rendelt intézmény tudomást szerezzen a dokumentumok létéről, ezeket megszerezze, tárolja, és azok rendelkezésére bocsássa, akiknek a dokumentumra (az abban foglalt információkra) szüksége/igénye van.** Ez az utóbbi két fogalom nem szükségképpen esik egybe, hiszen azok közül, akiknek ezekre az információkra szüksége van, csak azok igénylik/igényelhetik ezt, akik a dokumentum létéről tudnak, és arról olyan ismereteket szerezhetnek, amelyek alapján el tudják dönteni, hogy szükségük van-e az abban foglalt információkra. Mindez a könyvtári funkciót megtoldja egy újabbal: a dokumentumok létéről és főbb jegyeiről szóló információk rögzítésével és hozzáférhetővé (megtalálhatóvá) tételével, más szóval a **feldolgozás** vagy (ahogy a könyvtárosok némi rátartisággal és az előkelőbb közlésmód iránti vággyal nevezik), a **feltárás** funkciójával.

Jóllehet, e funkciók ellátásában más – egyre több fajtájú – intézmény (múzeumok, levéltárak, adatbázisok létrehozói/terjesztői, és – rohamosan bővülő mértékben – a közvetlen internetes kommunikációban résztvevők) osztoznak a könyvtárakkal, mégis ezek közül a legáltalánosabb és legelterjedtebb, a felsoroltakkal gyakran integrálódó, vagy azokat gyakran integráló intézmény: a **könyvtár**, helyesebben a más-más használói kört kiszolgáló és ennek megfelelően a dokumentumok más-más körét gyűjtő könyvtárak. Általános fontosságuk és elterjedtségük mellett a könyvtárak használóinak köre természetesen függ a könyvtár jellegétől, méretétől, nyilvánosságának fokától, kiszolgált olvasóinak társadalmi funkciójától, egyéni szándékaitól stb.

Az imént említett feldolgozás/feltárás tehát arra való, hogy a könyvtár használója **tudomást szerezzen** az érdeklődésére számot tartó (vagy azt felkeltő) dokumentumról, helyesebben e dokumentum jellegzetességeiről, mindenekelőtt az abban foglalt **információk tartalmáról és jellegéről**. A könyvtár, ha nem biztosítja ezt a lehetőséget, halottá, vagy legalábbis félholtta válik. **Ezért meghatározó a könyvtári funkciók közül az említett “tudomásszerzést” biztosító információkeresés, illetve az ezt szolgáló eszközök létrehozásának funkciója.**

Nem lehet egyértelmű jóslásokba bocsátkozni a **könyvtárak jövőjéről**. A könyvtárosok – és e munka írói (ha nem is eredetileg tanult szakmájuk szerint) könyvtárosok – más szakmák művelőiehez hasonlóan hajlamosak szakmájukat és az azt feltételező, annak otthont adó intézményt, vagyis a könyvtárat örökéletűnek feltételezni. Ha ezt a feltételezést indokoltnak tartjuk, annak ismét csak funkcionális okai vannak:

A feldolgozásról és az információkeresésről szólva már utaltunk rá, hogy a könyvtári funkció nem tölthető be **a könyvtáros segítő szerepe nélkül**. Az olvasói kör igényeit kielégítő, célzott könyvtári dokumentumbeszerzésen (vagyis a gyűjtőkör meghatározásán és érvényesítésén), a feldolgozáson, és az információkeresés eszközeinek előállításán túlmenően ez a segítség azt is jelenti, hogy a könyvtáros személyesen – e célból szerzett tudása birtokában – **eligazítja az olvasót** (könyvtárhasználót, információhasználót) a rendelkezésére bocsátott eszközök használatában.

Ha nem is tételezzük fel belátható időn belül (legalábbis addig, amíg ez az írás korszerűnek tekinthető) a nyomtatott dokumentumok számának és tartalmának abszolút visszaesését, a relatív visszaesés az **elektronikus dokumentumok előretörésével párhuzamosan kétségtelen**, különösen akkor, ha figyelembe vesszük az elektronikus dokumentumok körén belül a könyvtárak által még használati licenc formájában sem beszerzett, a **hálózaton rendelkezésre álló dokumentumok tényleges használatának túlsúlyát**. Vajon ez a fejlődés nem ingatja meg a könyvtári funkció olyan alapköveit, mint a gyűjtőkör szerinti állománygyarapítás, az állomány feldolgozása/feltárása, illetve a saját információkereső eszközök készítése és/vagy kezelése?

Úgy véljük, hogy nem. A hagyományos dokumentumok vélhetően még hosszú ideig tartó tömeges túlélése mellett az elektronikus információkat vagy azokat is használó olvasó is rászorul a könyvtár, a könyvtáros és a könyvtári információkeresés gyakorlatában létrehozott eszközeinek segítségére. A hálózati információk használatát azok közlői és közvetítői igyekeznek egyebek között az azonosítást és az információkeresést szolgáló – könyvtári hagyományoktól nem független – eszközök fejlesztésével segíteni. Ezért is, és ettől függetlenül is **a jövő könyvtárosának segítségét (aki mind saját könyvtári terében, mind a “kibertérben” otthonosan mozog) a jövő információhasználója, olvasója sem nélkülözheti**. Ez még akkor is igaz, ha eltekintünk számos olyan olvasási és más dokumentumhasználati céltől, amelyek esetében a hálózat és a közvetítő képernyő nem versenyképes a könyvtárral és a nyomtatott dokumentummal.

1.2 A táguló könyvtár

A könyvtár fogalmát az eddigiekben úgy írtuk le, mint olyan intézményét, amely a használói köre által vélelmezhetően szükséges dokumentumokat gyűjti, azokat a kiválasztás (más szóval: az információkeresés) biztosítására feldolgozza, és rendelkezésre bocsátja. Valójában már sok évszázados hagyománya van annak, hogy **a könyvtárak tevékenysége ehhez képest kibővült, tágult**. Ez a “tágulás” nem független az emberi tudás, a tudományok fejlődésétől, az írott kommunikáció mennyiségi fejlődésétől és differenciálódásától, a társadalomban végbement, az információhasználók körét és információigényeik intenzitását növelő és differenciáló fejlődéstől. Mindehhez a huszadik század második harmadától egyre gyorsuló ütemben csatlakoztak az információs technológia fejlődésének irányzatai, eredményei.

A könyvtári funkció első tágulását – ha nem is szükségszerűen mindig könyvtári keretekben – a **bibliográfiai tevékenység** jelentette, amelynek kezdetei az ókorig nyúlnak vissza, és így majdnem egyidős a könyvtárakkal, de amelynek itt említendő “tágító” hatása a késő középkorban és a korai újkorban (a bibliográfia atyjaként tisztelt Konrad Gesner tevékenységét követően) bontakozott ki

igazán. A bibliográfia – bár válogatási szempontjai igen eltérőek: földrajziak, nyelviak, történetiek és ezek összessége révén nemzetiek, tárgykörök, bizonyos dokumentumfajtákra orientáltak lehetnek – mindenképpen tágította a könyvtári horizontot, hiszen a bibliográfia jellegzetes vonása, hogy **mindenkori gyűjtőköre és gyűjtési forrásainak köre túllép az egyes könyvtári gyűjteményeken**. Ez alól adott történeti körülmények között – mutatis mutandis – kivételek a nemzeti bibliográfiák és nemzeti könyvtárak, de ez nem változtat az adott törvényszerűségeken. A bibliográfiák válogatási szempontjai és így összetételük adottságai alapján meghatározott információkeresési követelmények és ennek megfelelő eszközök jöttek létre. A könyvtár számára a bibliográfia – természetesen még a könyvtári szervezeti keretek közötti bibliográfiai tevékenység korántsem ritka jelensége is – azt jelentette, hogy növekedett az a tájékoztató/tájékozódási bázis, amire a könyvtár támaszkodhatott, és amely bázisból is képessé vált a kívánt információ keresésére vagy annak elősegítésére.

A bibliográfiai tevékenység fejlődésétől nem függetlenül **tágul az a dokumentumállomány** is amelyre a könyvtár támaszkodhat. Bár természetesen korábban sem volt ismeretlen, szükségképpen gyors fejlődésnek indul a könyvtárközi kölcsönzés és később – a technika fejlesztésével – a másolatszolgáltatás. Mind a bibliográfiai tevékenység, mind a dokumentumszolgáltatás és az annak feltételét képező információkeresés eszközeiként létrejönnek a könyvtári-bibliográfiai feldolgozás (ezen belül az információkeresésre szolgáló eszközök) nemzetközi szabványosításának törekvései, eszközei is.

Kissé visszalépve az időben találkozhatunk a “tágulás” további fontos jelenségével: a **dokumentáció (szakirodalmi tájékoztatás) kialakulásával** és jelentőségének ugrásszerű növekedésével. Mint ismeretes, e sajátos, bibliográfiai jellegű, de azt sok tekintetben megújító tevékenység ifjú szülőanyja a tudományos folyóirat volt. Az ok nem elsősorban a folyóiratok periodicitása volt, hanem az, hogy a folyóiratok “poligráfiai” jellege miatt a bennük foglalt művek (cikkek) a hagyományos, önálló művekre irányuló könyvtári feldolgozás eszközeivel nem voltak megragadhatók, nem is szólva arról, hogy valamely tudományág, illetve valamely téma művelőjének általában nem álltak rendelkezésére az idevágó források.

A bibliográfia (ez esetben a szakbibliográfia) és a tudományos cikk “szülőpárja” szinte a folyóirat mint forrás megjelenésével egyidejűleg hozta létre a folyóiratcikkek, tanulmányok tartalmát feltáró jellegzetes dokumentációs műfajt, a referátumot (csak azért írjuk ide angolul is, hogy “abstract”-et, mert sajnálatos módon a szaknyelv elangolosodása folytán már nálunk is sokan csak ezt a nevet ismerik) és a referátumokat közreadó referáló folyóiratot.

Mind a bibliográfiai, mind a dokumentációs/szakinformációs tevékenység fejlődése természetesen visszahatott az információkeresést lehetővé tevő feldolgozás, mindenekelőtt a tartalmi feltárás fejlődésére, az **osztályozó rendszerek, illetve a természetes nyelv eszközeit használó, szabványosított információkeresésre szolgáló** megoldások kibontakozására és nemzetközi elterjedésére.

A további tágulás – a mikromásolatok, mikrofilmek használatát és könnyű cseréjét, illetve az ezeket hasznosító információkereső eszközöket itt részletesebben nem is említve – már elvezet napjaink elektronikus korszakához. A számítógépes (offline, majd online) információkeresés tömeges méretekben azzal kezdődött, hogy elsősorban dokumentációs folyóiratokat (folyamatosan megjelenő szakbibliográfiákat és referáló kiadványokat) számítógéppel állítottak elő, és az így létrejövő, géppel olvasható adattárakat alkalmassá tették az információkeresésre. Kezdetben a használat formája a **mágnesszalagon** rögzített referáló folyóiratok vagy szakbibliográfiák információkeresésre alkalmas adatokkal ellátott másolatainak cseréje, (eladása/vétele) az ezek használatához szükséges információkereső programok kidolgozása és az ezek használatára alapozott információkeresés volt. Ez az eljárás – kezdetben elsősorban a természet- és alkalmazott tudományok területén – nagyban bővítette a könyvtárakban kereshető és a könyvtárak által kínálható információs választékot.

A már robbanásnak is nevezhető tágulás a **távoli forrásokban**, adatbázisokban, távközlési csatornákon át végzett információkeresés, az új információhordozókon (**CD-ROM**) **kereshetően**

rögzített adatbázisok beszerzése és használata, végül pedig a **saját adatbázisok létrehozása** és azokban végzett információkeresés eredménye. Az előre kiválasztott mágnesszalagos adattárakban végzett keresés folyamatába nem lehetett beavatkozni, legfeljebb meg lehetett ismételni a keresést más-más kereső szempontok megadásával, továbbá e mágnesszalagos információkeresés általában periodikusan közreadott adattárrészletekben volt elvégezhető. Az említett – lokális felhasználáshoz kapcsolódó – lehetőségekkel mindez megváltozott. Ilyen változás volt az, hogy egyre növekvő választékból kiválasztott **teljes adatbázisokban** lehetett keresni a kívánt információkat, és hogy esetenként lehetett megválasztani azt az adattárat, amelyből a kívánt információkat keresni akarták. Ennél is lényegesebb azonban, hogy a keresés folyamatát közvetlen beavatkozással módosítani lehetett, **online, interaktív információkeresésre** nyílt lehetőség. Ez a fejlődés visszahatott az információkeresés elméletére és gyakorlatára is, különösen eszköztárának fejlődésére.

A saját adatbázis-építés rövidesen meghódította a hagyományos könyvtári információkereső eszközök, a katalógusok világát is.

És végül az **internet**. Elvileg az internet is “csak” ugyanúgy tágítja a könyvtárakban, illetve a könyvtárak által végzett információkeresés határait, mint ahogy erről az online információkeresésről szólva már beszéltünk. Valójában azonban többről van szó. Szemben a korábbi, statikus gyakorlattal (a könyvtár és használója tudja, hogy hol és hogyan kereshet többé-kevésbé jól feltérképezhető adatbázisokban), az interneten a további keresésre alkalmas adatbázisok és maguk az elsődleges vagy annak álcázott, szintén kereshető **információk korábban elképzelhetetlen mennyiségben (és tegyük hozzá: teljes összevisszaságban) állnak rendelkezésre**. Az e célra szolgáló eszközök (azonosítók, metaadatok, keresőmotorok) ellenére az internet abban a korszakban van, hogy gazdagabb minden könyvtárnál, a könyvtári funkció gyűjtő és szolgáltató részének sohasem ismert bőségét valósítja meg, de rendetlenebb és – elsősorban a meglévő, de fel nem tárható információk tömege, a “beépített” információ-veszteség tekintetében – abszolút bőséget és relatív “elnyomorodást” hozott. Ennek a helyzetnek a megváltoztatásáért már sok minden történt, de még hosszú az út, amely az óriási lehetőségek adekvát használatához vezet.

A könyvtári funkciónak – ha fennmarad, és meggyőződésünk szerint fennmarad – világméretekben és könyvtáranként is megkülönböztetetten fontos feladata, hogy segítse az olvasót az internetnek, ennek a nagyszerű, de bonyolult eszköznek a használatában, ennek a nehezen megzabolázható táltosnak a megnyergelésében. Egy kissé emlékeztet a kép a fogyasztói társadalom “terülj asztalkáim”-ra, a bevásárló központokra. A választék óriási, de érteni kell ahhoz, hogy megtaláljuk, amit keresünk, a jelzőtáblák és feliratok pedig nem felelnek meg az igényeknek. Úgy véljük, hogy ennek eszköze – teljesebb fegyverzettel – a táguló és változó, de alapfunkciójában változatlan könyvtár.

1.3 Könyvtári információkeresés

Bár csak a következő fejezetekben igyekszünk betekintést nyújtani a könyvtári információkeresés különböző céljaiba, módszereibe, eszközeibe, lehetőségeibe és korlátaiba, indokoltnak látszik, hogy röviden már itt áttekintsük ennek a fogalomnak egyes fontosabb ismereti építőköveit.

Mindenekelőtt: a könyvtári információkeresés csak helyszínében, sőt, sok esetben még helyszínében sem, csak eredetében, fogantatásában, és környezetében, előképeiben különböztethető meg az **információkereséstől általában**.

Ha mégis könyvtári információkeresésről beszélünk, azt nem annyira ennek a könyvnek a címzett olvasóköre és megírásának célja indokolja, hanem egyfelől az eljárás és a név genezise, másrészt az, hogy (talán a levéltári funkció mellett, de ott is sokkal egységesebben) nincs olyan tevékenységi ág, amelynek funkcióiban az információk létrejöttékor általában **előre nem szabályozott jellemzők szerinti információkeresésnek és az információ kereséséhez szükséges feltételek megteremtésének olyan nagy súlya és jelentősége lenne, mint a – természetesen "tágult értelemben" felfogott – könyvtárakban**.

Itt jegyezzük meg, hogy az információkeresést sokan előszeretettel információ-visszakeresésnek nevezik. Ennek eredete, hogy a nyomkereső kutya (retrieval dog) kereső műveletének analógiája alapján elnevezett angol "information retrieval" és a francia "recherche" szó hangzása, különösen pedig a bevezető "re"- igekötő miatt sokan úgy vélik, hogy a magyar kifejezésnek is ki kell fejeznie: itt nem egyszerűen keresésről, hanem valaminek az "újra kinyeréséről", valami elveszettnek, elrejtettnek az újbóli megtalálásáról van szó. Valójában ez nem így van; (egyébként más nyelvekben sem). Ezt talán semmi sem bizonyítja jobban, mint az interneten végzett információkeresés. Amit ugyanis ott keresünk, azt biztos nem mi "tettük oda", nem visszakeressük, hanem keressük, megkeressük. Ezért maradunk e könyvben következetesen az "információkeresés", egyértelmű szövegkörnyezetben pedig a "keresés" szó mellett és ezt ajánljuk a gyakorlatban is. Hasonlóképpen nem követjük el azt a hibát, hogy az információról mindig egyes számban írunk az angol nyelv sajátos, logikátlan – de sajnos más nyelveket is fertőző – "egyesszám-szabálya" miatt.

Információkeresésről – és nem egyszerűen "keresésről" – pedig (hasonlóan az angol "information retrieval" kifejezéshez) azért beszélünk, mert ezzel jelezzük, hogy nem pusztán műveletről, hanem szakterületről van szó. Az információkeresés szakterületén belül – többek között – műveletek tárgyalhatók, mint például a keresés (searching), lekérdezés (scanning), a böngészés (browsing), szörfölés (surfing), vagy akár az innovatív keresés (serendipitous searching). Ezekkel a műveletekkel részletesen a későbbi fejezetekben foglalkozunk.

Az információt, információkat igénylő ember (a könyvtárban minden olvasó, minden használó ilyen ember) **valamilyen módon meghatározható, leírható információt keres**. Ebből a szempontból semmi különbség nincs a "valami jó szerelmezt tessék adni" mondattal, vagy az "etikával foglalkozó filozófiai munkákat keresek" kívánsággal jelentkező olvasó között; ezúttal még annyi sem, hogy két "mesebeli" olvasónk egyike sem meghatározott dokumentumot, még csak nem is egyetlen szerző munkáját, nem is valamely műfajba tartozó, esetleg speciális információhordozón rögzített dokumentumot kíván kikölcsönözni; mindkét olvasó igénye tárgyköri, tematikus.

A lehetséges **keresési szempontokat nem könnyű kategorizálni**. A könyvtári szakirodalomban hamarosan – kivált Európában, a doktriner törekvések hatására – két meglehetősen merev ágra szakadt az információkeresést és eszközeit megvalósító, illetve kutató szakemberek és az oktató diszciplínák, sőt az elméletek köre is. Míg a tárgyköri (tehát a könyvtári dokumentumok által tárgyalt) fogalmak, fogalomkörök szerinti információkeresés feltételeivel és az erre a célra szolgáló hagyományos vagy újabb eszközökkel az osztályozók és/vagy indexelők (a magyar szaknyelv szörnyű tolvajnyelvi szava szerint: a "szakozók") foglalkoztak, addig az egyéb szempontok szerinti információkeresés feltételeivel, eszközeivel a katalógizálók.

A gyakorlatban **nincs áthidalhatatlan szakadék a különféle keresési szempontok és a különféle keresési igényekre választ adó eszközök között**. Így például a hagyományos megközelítés szerint az egyik kategóriába, nevezetesen a (szerző, vagy közreműködő) személyek neve szerinti kategóriába tartozó információkeresés eszközeinek megteremtése a katalógizálás ("formai feltárás") feladata. Az osztályozási jelzetek, tárgyszavak, deskriptorok hozzárendelése pedig a "tartalmi feltárás" feladata. Könnyű belátni azonban, ha például valamely személyről szóló mű (dokumentum) kereséséről van szó, akkor tárgykör szerinti keresést végzünk, de hozzá az egységesített személynevet kell használnunk, és ennek feltételeit ugyancsak a katalógus – tehát a katalógus kialakítása, szerkesztése, azaz a katalógizálás – segítségével kell megteremteni. Hasonló példák sokaságát említhetnénk még.

Természetesen ez nem jelenti azt, hogy a könyvtári dokumentumokról készülő, az információkeresés eszközt jelentő leírások minden elemének azonos a fő funkciója. Az információkeresésre szolgáló eszközök körét a felmerülő keresési igények és a meghatározható keresési eszközök lehetőségei, adottságai szerint kategorizálni kellett. A lehetséges keresési pontok meghatározásához, kiválasztásához és leírásukhoz különböző módszerek és eszközök kellettek; a keresés szempontjaként felmerülő fogalmak és nyelvi megfelelőik különböző mértékben **egységesítettek vagy akár szabályozottak**.

Ismételten hangsúlyozva, hogy a keresés szempontjait és így eszközeit nem választja el egymástól valamiféle kínai fal, a gyakorlatban (és ebben a könyvben is) mégis különbséget kell tennünk

egyfelől a könyvtárban tárolt információk/dokumentumok tárgyköre, másfelől egyéb jellemzői (létrehozó személyek, testületek, a dokumentumon található, azonosításra alkalmas szövegek, a dokumentum megjelenésének adatai, fizikai jellemzői, más dokumentumokkal való rokonsága, azonosítási célra létrehozott adatok stb.) között. Ez utóbbiakat szokták összefoglalóan formai jellemzőknek is nevezni, szemben az előbbi, tartalmi jellemzőkkel. E kétfélének tartott jellemző feldolgozása a munkamegosztás következtében gyakran (de nem mindig és nem feltétlenül) elkülönül egymástól. Az információkeresés területén azonban ilyen munkamegosztásnak nem volt gyakorlati értelme és haszna.

Az információkeresés **nem merül ki a keresési pontok/ismérvek meghatározásában**. A keresési ismérvek rögzítésének, elrendezésének, hozzáférhetővé tételének, és közöttük fennálló összefüggések kimutatásának különböző útjai és módjai vannak. Hasonlóan eltérő lehet – amint erre már utaltunk – a keresés technikája, technikai eszköze, lebonyolítása, folyamata is. E könyv igyekszik ezeket az eszközöket és folyamatokat is áttekinteni.

Bár (talán némi önzéssel, de úgy véljük indokoltan) a könyvtárakat helyeztük az információkeresés világának középpontjába, természetesen más tevékenységek során, sőt mindennemű, nagymennyiségű és változatos információt tároló és kezelő rendszerben és tevékenység körében is felmerül az információkeresés szükségessége. Erre a technikai fejlődés korábbi szakaszaiban is rengeteg példa volt, nem is szólva arról, hogy az írott dokumentum, tehát a könyvtári tevékenység tárgyának alkotói, amikor tartalomjegyzéket és mutatót alkotnak, maguk is az információkeresés eszközeit hozzák létre. E tekintetben két szembeötlő fejleménnyel kell számolni. Egyrészt az automatikus indexelés és osztályozás lassú kialakulásával. Ez a folyamat valójában még ki sem bontakozott igazán, de jövőbeli jelentősége kétségtelen. E teljesen automatizált eljárások azonban ugyanúgy nem fogják szükségtelenné tenni az intellektuális információfeldolgozást és keresést, ahogy a gépi fordítás sem képes kiváltani a szellemi munkával végzett fordítást. Mind a teljesen automatizált, mind a szellemi eljárásoknak megvan és meglesz a maguk helye. Másrészt számolnunk kell – és ez ma a legszembeötlőbb fejlemény – az internet létevel és az **interneten végzendő információkeresés kihívásával**. A gyakorlati tapasztalatokhoz igazodva és a lehetséges elvárásokat mérlegelve úgy véljük, hogy – minden nagy jelentőségű különbség ellenére – az internetes információkeresés is könnyebben megoldható feladat mindazon eszközök használatával, amit a tág értelemben vett könyvtári információkeresés kitermelt.

Az eddigiek összefoglalásaként arra kérjük az olvasót, hogy a részletesebb tudnivalókat tartalmazó fejezetek tanulmányozásakor ne feledkezzenek meg azokról a sokrétű összefüggésekről, amelyeket az eddigiekben vázolni próbáltunk.

2. Az információkereső nyelv és szótára

2.1 Miért kell sajátos nyelv az információkereséshez?

Könnyű belátni, hogy ha valamilyen információkereső kérdést azokkal a szavakkal fogalmaznánk meg, melyek mindennapi vagy szakmai nyelvhasználatunk alapján az eszünkbe jutnak, akkor legfeljebb abban az esetben lenne reményünk sikerre, vagyis hogy megtaláljuk azt, amit keresünk, ha szakavatott segítő állna rendelkezésünkre, aki egyrészt megértené kérdésünket, másrészt olyan készségekkel, képességekkel, eszközökkel rendelkezne, amelyek lehetővé tennék számára, hogy megtalálja számunkra azt, amit keresünk. Ebben az esetben azonban csak áthárítottuk a keresés feladatát, és elodáztuk azt, hogy kifejlesszük azokat az eszközöket, amelyek segítségével a keresett “valamit” megtalálhatjuk. A megoldáshoz megfelelő rendszerre és annak eszköztárára van szükségünk. Ezek egyike az **információkereső nyelv**. Az alábbiakban – anélkül, hogy részleteket tárgyalnánk a különféle információkereső nyelvekről és használatukról – áttekintést adunk e nyelvek szükségességéről, jellegéről, fajtáiról. A különféle információkereső nyelvekre vonatkozó részletek tekintetében a megadott irodalomra utalunk.

Hogyan is keressük valójában azt, amit meg akarunk találni – nem csak információt, hanem bármit?

Abból indulunk ki, hogy igényünk van “valamire” – mindegy, hogy miért. Másrészt rendelkezésünkre állnak különféle “dolgok” és “egyedek” halmazai: emberek, könyvek, hangok, színek, ábrázolások és bármely más kereshető dolgok halmazai, amelyek esélyt kínálnak arra, hogy az igényelt (keresett) “valamit” az adott halmazban megkeressük, megtaláljuk.

A keresés maga úgy megy végbe, hogy a rendelkezésre álló halmaz elemeit **összehasonlítjuk** (mi magunk vagy – bonyolultabb esetben – egy erre hivatott rendszer összehasonlítja) a keresett egyedről alkotott, vagy esetleg rendelkezésünkre álló “képpel”. Gondoljunk csak arra, hogy ha kisebb-nagyobb embercsoportban keresünk valakit, akkor az illetőről tudatunkban lévő, kialakult vagy tudomásunkra jutott “képünket” vagy a “személyleírását” ismertető jegyeket hasonlítjuk össze a halmaz – ezúttal egy embercsoport – megfelelő elemeivel, mindaddig, amíg e “képet” valakinek a látványával, pontosabban annak érzékelt elemeivel azonosnak vagy azonossággal biztatónak, hasonlatosnak találjuk. Ha az a csoport (halmaz), amelyben keresünk, nem nagyon nagy, és a halmazt alkotó egyedek viszonylag jól meghatározható, “szembeötlő” tulajdonságokkal rendelkeznek, a halmaz adottságainak megfelelő keresési igény esetén a keresés (“találathoz” jutás) nem is nehéz.

A kívánt “kép” és a képnek megfelelő, illetve ilyenek talált egyed(ek) összehasonlított elemei az információkeresés során használt **ismérvek**.

Amikor valamely halmaz nem külsődleges tulajdonságokkal rendelkező elemei közül kell “választani”, különösen pedig ha a keresett és megtalálható elemeknek száma nagy, összetétele pedig bonyolult, akkor valamilyen **segédeszközhöz kell folyamodnunk**: először is rögzítenünk kell a keresett egyedeknek (a keresés tárgyának) tulajdonságait, ismérveit, hogy ezeket összehasonlíthassuk a keresett egyed(ek)et feltételezhetően magában foglaló halmaz egyedeinek tulajdonságaival, ismérveivel. Ezek korántsem mindenkor állnak “maguktól” rendelkezésre. Ha netán saját könyvespolcunkon keressük egy adott szerző adott könyvét, akkor a könyv gerincén megtalálhatjuk azt a feliratot, amely a keresett könyv szerzőjét és címét írja le, vagy legalább ezek közül az egyiket. Ez esetben tehát bizonyos ismérvek a könyvön rögzítve voltak, és összehasonlíthatók a keresett könyv szerzőjének nevével és a könyv címével. Általában azonban nem ez a helyzet, hanem az, hogy a keresés céljára kell rögzítenünk kérdésünk ismérveit, és olyan halmazban kell keresnünk, amely szintén – valamilyen módon rögzítve – tartalmazza, vagy amelyben mi rögzítjük a halmaz egyedeinek ismérveit.

Más szóval: **le kell írunk** a keresett egyed, illetve a keresett egyed(ek) remélt “lelőhelyét” képező halmaz elemeinek tulajdonságait, ismérveit. Ezt látszólag cáfolni lehetne azzal, hogy erre nincs szükség a keresett egyedtel való tartalmú halmaz oldaláról, ha ez a halmaz eleve olyan szerkezetű, hogy ez közvetlenül lehetővé teszi a keresett egyed leírt tulajdonságaival való összehasonlítását. Ha azonban meggondoljuk ezt az ellenvetést, azt láthatjuk, hogy nem állja meg a helyét, hiszen az említett szerkezet, vagyis a halmaz egyedeinek besorolása valamilyen rendszerbe szintén az egyedek tulajdonságainak egyfajta leírását jelenti. E könyv szövegekörnyezetéből adódóan az “egyedek”, amelyek kereséséről szó van, információk (a magyar nyelvben – szemben az angol nyelv minden logikát nélkülöző szabályával – az információ szónak van többes száma), következésképpen az információkereső kérdéseket és a keresés tárgyát képező információhalmaz elemeit (amelyek szintén valamely dolog, folyamat, állapot, helyzet, esemény, elmélet, törvény stb. leírását jelentik) kell leírni. Ahhoz, hogy valamit leírjunk, tudatunk szerkezetéből adódóan a fogalmak manifesztációjára, vagyis valamilyen nyelvre van szükségünk. Azt a nyelvet, amely alkalmas az információk és az információkeresési kérdések leírására, vagyis a természetes nyelven megfogalmazott leírások egységes nyelvre fordítására, **információkereső nyelvnek** nevezzük. E nyelvnek – mint minden nyelvnek – funkciójából adódó követelményeket kell teljesítenie.

Mondanunk sem kell, hogy az információkereső nyelvek főbb kérdéseit ebben a munkában elsősorban könyvtári szempontból vizsgáljuk. Ez azt jelenti, hogy esetünkben a keresőkérdés az olvasó, az információhasználó vagy az őt kiszolgáló, neki segítő könyvtáros kérdése, a keresett információ pedig a kereső igényeinek megfelelő tartalmi és/vagy egyéb adottságokkal rendelkező **dokumentum vagy a dokumentumban foglalt információ** (természetesen függetlenül attól, hogy milyen (hagyományos vagy elektronikus) dokumentumról van szó).

2.2 Melyek az információkereső nyelvvel szemben támasztandó követelmények?

Az információkereső nyelvvel szemben támasztandó első követelmény, hogy a keresés jellemzőit (a továbbiakban: a kérdést) és a keresés tárgyát “rejtő” halmaz egyedeit ugyanazon a nyelven kell leírni (ez a követelmény önmagában kevésbé sajátos: a kommunikáció általában azonos nyelv használatát tételezi fel az adó és a vevő oldalán). Másképpen a keresés, vagyis az összehasonlítás nem lenne lehetséges. Könnyű belátni, hogy ha a “fiatalember” kifejezést hasonlítjuk össze a “jeune homme” kifejezéssel, akkor e két ismérv azonosságát nem lehet megállapítani. Úgy tűnhet, hogy ha a keresést franciául tudó ember, vagy francia–magyar szótárt alkalmazó gép végzi, akkor ez lehetséges lenne, valójában azonban akkor is egyetlen nyelven, a jeune homme-ot fiatalemberre, vagy a fiatalembert jeune homme-ra fordítva állapítható meg az azonosság. Az információkereső nyelvnek tehát **egyetlen nyelvnek kell lennie**.

Egyetlen nyelv azonban még nem jelent **egyértelmű** nyelvet. A hétköznapi vagy akár a tudományos gyakorlatban használt nyelv – nevezzük az információkereső nyelvtől való megkülönböztetés végett természetes nyelvnek – szavai, kifejezései két tulajdonságukból adódóan is vétenek az egyértelműség követelménye ellen. Egyik ilyen tulajdonság az, hogy ugyanannak a szónak vagy kifejezésnek több jelentése (nem szükségszerűen rokonértelme) lehet: elegendő példaként csak olyan egyszerű szavakat említeni, mint: nap, ár, kelet, fűző stb. (de még sok ilyen példát lehetne felsorolni). Ezt még gyakoribbá teszi számos kifejezés tolvajnyelvi jelentése (például ruha, lé, sitt). E jelenséget (kissé egyszerűsítve) **homonímiának, poliszémiának** vagy – a hangzást helyezve előtérbe **polifóniának** nevezik. Nem kell külön bizonyítani, hogy az információkereső nyelv e szavakat nem, helyesebben csak értelmezve használhatja.

Ugyanezen jelenség ellentettje az eltérő írásmódú, illetve hangzású, de azonos, vagy igen hasonló jelentésű szavak, kifejezések léte. Tágabb értelemben ide sorolandók az azonos hangzású szavak eltérő írásmódú alakjai is, de korántsem csak erről van szó. Részegség és ittasság, ródli és szánkó, irka és fűzet, kutya és eb ugyanazt jelentő, de teljesen eltérő írásmódú és hangzású szavak. Egyáltalán nem csak a hétköznapi nyelvben, hanem gyakran a tudomány, illetve a különböző tudományok területén is találkozunk ezzel a jelenséggel. Mint ismeretes, itt a **szinonimákról** (például eb és kutya) van szó, amelyek népes családjának “oldalági” rokonai a kváziszinonimák (például fű és gyep, alkohol és etilalkohol¹, vagy hegycsúcs és hegyorom). De ide sorolhatók az idegen nyelvből egy adott nyelvbe átvett szavak (például számítógép és komputer), valamint a fő jelentésük mellett valamely más – kezdetben gyakran tolvajnyelvi – értelemmel is rendelkező szavak, kifejezések (például gépkocsi, kocsi, autó).

Talán nem kell bővebb magyarázat ahhoz, hogy ha az információkereső nyelv (illetve alkotója és használója) nem találna megoldást a poliszémia jelenségéből adódó nehézségek kiküszöbölésére, akkor e jelenség következtében a keresés téves eredményekkel (az információkeresés elméletében és gyakorlatában zajnak nevezett jelenséggel) járna, a szinonímia pedig a kérdés tárgyában előforduló, a kérdésnek megfelelő, de meg nem talált egyedek előfordulására (veszteségre) vezetne. Mivel pedig az információkereső nyelvnek – mint más nyelveknek – szótára van, az említett megoldás(ok)nak be kell épülniük az információkereső nyelv szótárába is, általában magyarázatok illetve utalások formájában. Az információkereső nyelvek szótáraiba felvett szavakat lexikai egységnek nevezik.

Mint azonban látni fogjuk, az egyetlen, egyértelmű nyelv sem képes megbirkózni mindazokkal a feladatokkal, amelyek az információkeresés feltételeinek megteremtése, előkészítése és végrehajtása során jelentkeznek.

2.3 Az értelmi (tartalmi, jelentésbeli, paradigmaticus) összefüggések

¹ Az “alkoholos” példa azt szemlélteti, hogy – elsősorban a köznyelvben – gyakran bővebb értelmű (általánosabb) és szűkebb értelmű (speciális) szavak, kifejezések lépnek fel kváziszinonimaként.

További követelményeket is kell támasztani az információkereső nyelvvel szemben. A mindennapi életben, illetve a keresés egyszerű eseteiben az emberi ismeretek (jóllehet számítógéppel segített, különösen pedig automatizált megoldások esetén ezek és alkalmazásuk modellezése egyáltalán nem egyszerű) gyakran áthidalják a szavak, kifejezések, illetve az általuk kifejezett fogalmak összefüggései által okozott információkeresési nehézségeket. Ez azonban az információkereső rendszerekben – különösen nyelvi segédeszközök nélkül – nem általánosan alkalmazható és nem is megbízható megoldás.

Ha valaki háziállatokat (vagy velük foglalkozó dokumentumokat) keres, az információkereső nyelv pedig nem mutatja ki és nem teszi nyomon követhetővé, hogy a háziállatok fogalmi körébe tartoznak a különféle “saját nevükre hallgató” (ami alatt persze nem Bodri vagy Ráró értendő) háziállatok, vagyis a háziállatok fajtái (kutya, macska, sertés, ló, szarvasmarha, kecske, juh, különféle baromfiak stb.), akkor információvesztés keletkezik, hiszen a “háziállat” szó összehasonlítása például a “kutya” szóval nyilván negatív eredményt ad. Az információkereső nyelv szótárának ezért **ki kell fejeznie a bővebb, általánosabb (nem-) és a szűkebb, specifikus (faj-) fogalom közötti értelmi (tartalmi) összefüggéseket**, (kapcsolatokat, viszonyokat, relációkat) mindkét irányban.

Az említett jelentésbeli összefüggések egyedi – de korántsem egyedüli – esete a fenti példa szerinti bővebb és szűkebb értelmű, általánosabb és speciálisabb jelentésű kifejezések (fogalmak) összefüggése.

Hasonló a helyzet az **egész és részei** (például vérkeringési rendszer, illetve szív, erek, vér) esetében, de a különbség is jól érzékelhető: itt nem bővebb és szűkebb fogalmakról van szó, hanem az egészet és alkotórészeit kifejező fogalmakról (illetve azokat reprezentáló szavakról, kifejezésekről). Mégis: akit a keringési rendszer érdekel, azt nyilván érdeklik az említett “rész-fogalmak” is, akit pedig az erek érdekelnek, annak “válaszolhatnak” a keringési rendszer egészéről szóló információk is.

Folytatva a sort: aki az oktatás iránt tudakozódik, annak érdeklődésére nyilván számot tart az, ami az iskolával, az egyetemmel, a tanárral, a tantervvel, a tankönyvvel, szemléltető eszközökkel és más, e “rokonságba” tartozó fogalmakkal kapcsolatos. Itt arról van szó, hogy szoros értelmi összefüggés áll fenn egy **cél vagy rendeltetés és az elérésére, megvalósítására szolgáló eszközök vagy egyéb, oksági kapcsolatban álló dolgok** között.

Még számos összefüggésfajtát kell vagy lehet az információkereső nyelv szótárának – mint látni fogjuk – egyedileg vagy összevontabban, különféle lehetséges eszközökkel kifejeznie. Valójában a szinonim szavak (kifejezések) között is értelmi, jelentésbeli összefüggés (reláció) áll fenn: méghozzá az, hogy a szinonim kifejezések (szavak) azonos vagy igen hasonló értelműek, jelentésűek.

Mivel ezek az összefüggések a szótárban előre meghatározott formában és sorrendben – tehát előre megadott minta szerint – szerepelnek, a nyelvészetből kölcsönzött kifejezéssel nevezik őket paradigmikus relációknak is (a paradigma nyelvi, általánosabban kulturális mintát, szabályt, példát jelent).

2.4 Az adott “esettől” (szövegkörnyezettől) függő összefüggések

Még az egyértelmű, a fogalmak közötti, jelentésbeli összefüggéseket kimutató információkereső nyelv sem feltétlenül vagy legalábbis nem tökéletesen alkalmas a kereső által támasztott igények, illetve a keresett halmaz elemei közötti összehasonlítás céljaira. A megfogalmazott keresési igény (keresőkérdés) és a keresés leírt tárgyának elemei, szavai, **esetenként eltérő összefüggésben** állhatnak egymással, ahogy egy mondatot alkotó szavak és kifejezések is különböző összefüggésben állhatnak. Az ilyen összefüggések nem következnek az információkereső nyelv szavaiból, hanem – a kereső igényétől, illetve a keresett egyed adottságaiból következően – azok adott használatát határozzák meg. A természetes nyelv mondatai ezeket az összefüggéseket adott szövegben különböző szófajok alkalmazásával, szavak toldalékaival, elöljárókkal (ha nem is a magyar nyelvben), szórenddel és más nyelvi eszközökkel fejezik ki. Ezért ezeket az

összefüggéseket **mondattani, szintaktikai összefüggéseknek** is szokták nevezni. Kívánatos lenne, hogy az információkereső nyelv is képes legyen ezeknek az adott esetben érvényes szintaktikai összefüggéseknek a kifejezésére, vagyis erre alkalmas eszközökkel rendelkezzen. A nem gépi információkeresés céljaira előállított eszközökben (például különféle speciális mutatókban) voltak és vannak erre szolgáló eszközök, de a gyakorlatban alkalmazott információkereső rendszerekben – legyenek azok akár hagyományos katalógusok vagy mutatók, akár online számítógépes információkereső rendszerek – ilyenek használata csak igen ritkán, kivételesen fordul elő. A keresőkérdések összeállításakor használatosak a kereső fogalmak közötti főbb összefüggéseket kifejező eszközök, de nem a keresés tárgyát képező, információkereső nyelven leírt fogalmak egymás mellé rendelésekor.

Így aztán szükségszerű, hogy – leegyszerűsített példán bemutatva a következményt – ha valaki a programozás oktatásával kapcsolatos információkat keres a “programozás” és az “oktatás” fogalmak információkereső nyelvi megfelelőjével, olyan eredményt is kaphat, amelynek esetében nem erről, hanem programozott oktatásról van szó. Ez a példa is mutatja, hogy az összetettebb információkereső nyelvi kifejezések (például “programozott oktatás”) alkalmazása csökkentheti a szintaktikai összefüggések elhanyagolásának káros hatását és az emiatt előálló zajt, másfelől azonban az összetett kifejezések használata bizonyos határon túl hasonlíthatatlanul bonyolultabbá és kevésbé működőképpé tenné az információkereső nyelv szótárát, illetve a keresés eszköztárát. A gyakorlatban azt a kompromisszumot szokták megkötni, hogy az adott gyűjtőkörben gyakrabban használt, vagy az értelmi összefüggések rendszerében fontos, csomóponti szerepet játszó esetekben, továbbá ha az összetételből nem állapíthatók meg egyértelműen az összetevők (például vitaülés), felveszik az összetett kifejezéseket, más esetben pedig nem.

Itt utalunk még arra, hogy amikor a keresés nem tárgykör szerint, hanem például személyek vagy testületek neve szerint folyik, akkor az információkereső kérdés nem a fentiekben jellemzett keresőnyelvi szabályokhoz igazodik, hanem tényleges nevek egyértelmű azonosítására törekszik. Hasonló – kiegészítő, illetve átmeneti – megoldások lehetségesek olyan ismeretágak területén is, amelyeknek általánosan elfogadott megnevezésrendszere (taxonómiája) van, de ez utóbbi esetekben az értelmi (jelentésbeli) összefüggéseknek ismét számottevő jelentősége van.

2.5 Az információkereső nyelv elemei

Az eddigiekben említett követelmények teljesítése, illetve teljesíthetősége és – ezzel együtt – az információkereső nyelv szavainak jellege, illetve az ezek szótárát alkalmazó mondatok (keresőképek) szerkezete elsősorban az információkereső nyelv szerkezetétől és az ezt nagyrészt meghatározó alkotóelemeitől függ. Mint minden nyelvnek, az információkereső nyelvnek is a **jel** a legkisebb eleme, ami persze lehet hang is, de az ebben a szöveggörnyezetben elsősorban jelentőséggel bíró írott nyelvek esetében általában betű (esetleg szó- vagy szótagjel), számjegy vagy írásjel (amelyek mindegyike gépi alkalmazás esetén természetesen bináris alakban is megjelenhet). Azt gondolhatnánk, hogy a jelkészlet minemúsége kevésbé jelentős az információkereső nyelv minősége, használhatósága szempontjából. E feltételezés nincs is nagy távolságra a valóságtól, amennyiben a korábbiakban vázolt követelmények elvben az alkalmazott jelek jellegétől, mibenlététől függetlenül teljesíthetők lehetnek. Teljesítésük módjára és különösen az információkereső nyelvek szótárának szerkezetére, ennek folytán pedig az értelmi összefüggések kimutatásának lehetőségére azonban már jelentős hatást gyakorolhat a **jelkészlet**. E hatásokkal foglalkozik/foglalkozhat az információkereső nyelvek jelelmélete, **szemiotikája**.

A betűk és más jelek különböző módon rendelhetők egymáshoz, vagyis belőlük különféle módokon lehet **szavakat képezni**. Az egyik lehetőség az, hogy ezek egymáshoz rendelése valamelyik **természetes nyelv szabályait, helyesebben gyakorlatát követi**. Ez esetben a jelek egymás mellé rendelésének módja – legalábbis az esetek túlnyomó többségében – nem fejez ki semmit, pontosabban semmi mást, mint a természetes nyelv adott kifejezése. Ha azonban nem ezt az utat követjük, hanem például az 1 számjeggyel jelöljük azt a fogalmat, hogy ember, illetve a 11 számmal azt, hogy férfi, a 12 számmal pedig azt, hogy nő, akkor itt az egymáshoz rendelésnek

értelme van: a második számjegy az első számjeggyel együtt hordozza azt az értelmet, amit az 1 számjegynek adtunk (ember), második számjegye pedig – az első számjeggyel együtt – a két nemet, tehát az ember fogalmának két változatát, alárendelt fogalmát jelzi. A példában szereplő számok helyett betűk is állhatnának, például A az ember. AA a férfi és AB a nő fogalmának kifejezésére. Az azonban nyilvánvaló, hogy ha az információkereső nyelv jeleiből értelmi összefüggések kifejezésére alkalmas módon képezünk jelcsoportokat (szavakat), akkor sajátos szóképzést alkalmaztunk, és ez befolyásolja nyelvünk minőségét, ami a **fogalmak közötti értelmi (jelentésbeli) összefüggések kimutatásának módját** illeti. Ez azt jelenti, hogy a szóképzésnek is korlátozó vagy elősegítő szerepe van illetve lehet az információkereső nyelvek **jelentésben, szemantikájában**. Látni fogjunk majd, hogy ez a rokonszenves lehetőség veszélyeket is rejt magában. Megjegyezhető azonban, hogy ezt a fajta szóképzést **hierarchikusnak is nevezzük**, hiszen a példánkban szereplő kétbetűs “szavak”, illetve az azokkal kifejezett fogalmak **fajtái, alárendelt fogalmai** az egy betűvel megjelölt szónak (fogalomnak). A szóképzés módja tehát nagymértékben befolyásolhatja az információkereső nyelv minőségét. Itt említjük meg azt is, hogy a “szóképzésnél” alkalmazott különféle jelek is értelmet, illetve értelmi összefüggéseket fejezhetnek ki.

Az értelmi, jelentésbeli, **szemantikai összefüggések kifejezésének azonban más útjai** is vannak, amelyeket főként akkor alkalmazunk, ha a szóképzés a természetes nyelv szavait követte, vagy – egyszerűbben kifejezve – **ha az információkereső nyelv a természetes nyelvből kölcsönzi “szóképzési szabályait”**. Ezek az utak és módok nemcsak minden könyvtárosnak, hanem minden könyvtárlátogatónak, sőt többé-kevésbé minden írni-olvasni tudó embernek ismerősek. Ide tartoznak a különféle **utalások és magyarázatok**. Egyes értelmi összefüggések, például az azonos értelmű (szinonim) szavak, kifejezések azonos értelműségének jelzésére más módszereket általában nem is alkalmaznak, jóllehet az utalások jelekkel is helyettesíthetők lehetnek, például az = jelet használhatjuk annak kifejezésére, hogy A és B szinonimák ($A = B$). A későbbiekben említendő természetes nyelven alapuló (szemantikai, analitikus) információkereső nyelvek igen differenciált jelölésrendszerrel (általában az értelmi összefüggés jellegére utaló betűvel) fejezik ki a szavaik közötti jelentésbeli összefüggés jellegét.

Végül megvizsgálandó, hogy milyen eszközöket használnak (és használnak-e egyáltalán eszközöket) az információkereső nyelvek az eseti, adott “szövegkörnyezeten” belüli **szintaktikai összefüggések kifejezésére**. A keresést szolgáló szövegek, mutatók alkalmazásakor voltak ilyen törekvések, de ezek az információkeresés tömegessé válásával és online megvalósításával “kimentek a divatból”. Ugyanakkor az információkereső nyelvre való automatikus fordítás meghozhatja e törekvések bizonyos reneszánszát.

Mielőtt rátérnénk az információkereső nyelvek és szótáraik főbb fajtáinak vázlatos bemutatására, előre kell bocsátani egy általános arany szabályt. Visszatérő betegség könyvtárosok köreiben, hogy mintegy “versenyeztetik” a különféle információkereső nyelveket, vagyis el akarják dönteni, hogy melyik fajtájuk a “legjobb”. Valójában a különféle nyelvi struktúráknak megvannak a maguk előnyei és hátrányai. Ezért adott esetben hasznos megoldás lehet többféle struktúra egymás melletti párhuzamos alkalmazása (mindenekelőtt a később említendő hierarchikus információkereső nyelveké és szemantikai információkereső nyelveké), és igen rossz megoldás akár a szubjektíve – például egyszerűsége miatt – kedvelt típusnak az előnyök és hátrányok voltaképpen vizsgálata nélküli fetisizálása. Az információkereső nyelvek különböző fajtáit csak vázlatosan mutatjuk be, elsősorban fő sajátosságaik, előnyeik és hátrányaik jellemzésére. A részletesebb tudnivalók tekintetében a megadott irodalomra utalunk.

2.6 Hierarchikus információkereső nyelvek

Említettük, hogy egyes információkereső nyelvek a **fogalmak közötti értelmi, jelentésbeli összefüggések** és ezen belül is – legalábbis elvben – mindenekelőtt az alárendeltség (tehát az általánosabb és speciálisabb fogalmak összefüggése) **alapján képezik a nyelv szavait**. Ez azt

jelenti, hogy e “szavaknak” általában nincs közvetlen, természetes nyelvi értelme hanem ezek a fölérendelt fogalom (szintén nem természetes nyelvi jelentésű) “nevéből” és az ahhoz hozzáadott megkülönböztető elemekből jönnek létre. E szóképzés azt jelenti, hogy ezeknek az információkereső nyelveknek minden szava “hurcolja magával” a neki fölérendelt fogalmakat. Ezt a megoldást alkalmazzák többnyire, vagy legalábbis gyakran az **osztályozási rendszerek**, ahogy e hierarchikus információkereső nyelveket a gyakorlatban nevezik². A szavaikat a természetes nyelvből kölcsönző információkereső nyelvek természetesen szintén kifejezhetik a fogalmak közötti hierarchikus összefüggéseket, de az itt tárgyalt értelemben nem hierarchikusak, mivel “szavaik” a hierarchikus összefüggésektől – általában – függetlenül épülnek fel. Nem sorolhatók a szűkebb értelemben vett hierarchikus információkereső nyelvek közé azok az osztályozási rendszerek sem, amelyek szavaikat kizárólag a természetes nyelvből kölcsönzik, de amelyek szótárai minden szavukat az alá/fölérendeltség szerkezetében tartalmazzák. A természetes nyelveken alapuló, mellérendelő információkereső nyelvek szavaiból többnyire nem “látszik”, hogy az illető szó, az általa kifejezett fogalom (vagyis a szó jelentése), milyen fogalmaknak alárendelt, következésképpen ezek nem hierarchikus információkereső nyelvek, még akkor sem, ha kimutatják a fogalmak hierarchikus összefüggéseit. (Léteznek azonban természetes nyelven alapuló nem mellérendelő, hanem hierarchikus osztályozási rendszerek is.)

A következőkben csupán azokat az információkereső nyelveket tekintjük hierarchikusnak, ahol a nyelv szavai implicite hordozzák a hierarchiát, más szóval az információkereső nyelvi szavakból kiderül, hogy az általuk reprezentált fogalom mely másik fogalomnak specifikus fogalma, vagyis melyeknek van alárendelve.

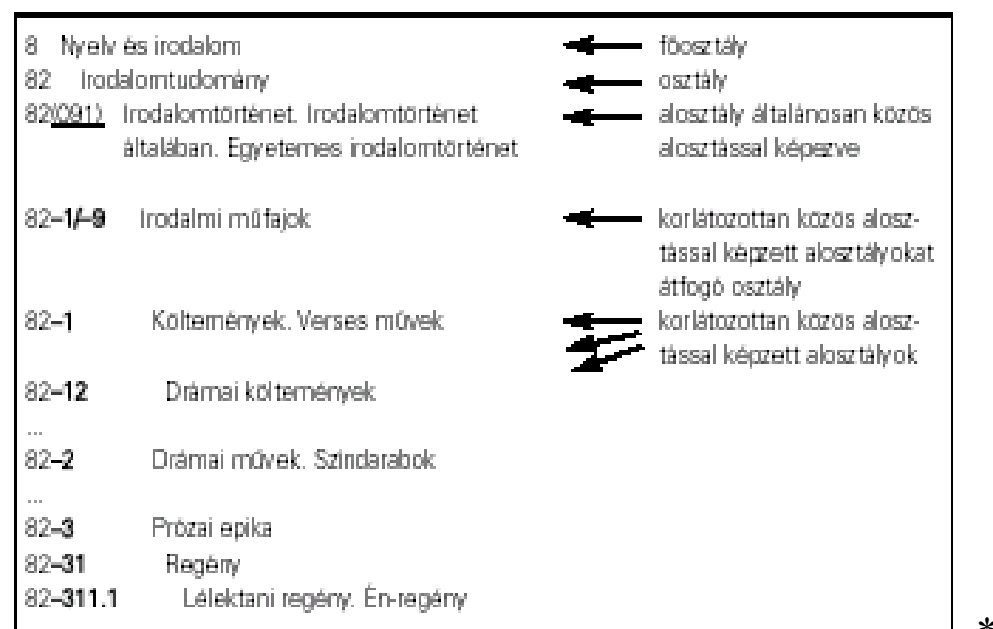
A hierarchikus információkereső nyelvek mindenkor teljes tematikájukat egyetlen egészként fogják fel. Ez a tematika különféle szélességű és mélységű lehet, mint például az emberi ismeretek egyeteme az Egyetemes Tizedes Osztályozás (ETO) vagy elődje, Dewey Tizedes Osztályozása (TO) esetében, vagy bizonyos ismeretágak, esetleg tevékenységi ágak különböző szakterületi (például fizikai, matematikai, vagy éppenséggel kohászati) osztályozó rendszerei esetében. Közös e rendszerekben, hogy az említett **“egész” (fölérendelt) fogalmat meghatározott szempont szerint “osztják fel”** alárendelt fogalmakra, és e szempont szerinti, annak megfelelő jelentésű megkülönböztető jegyeket rendelnek az említett egészhez (korábbi, kezdetleges példánkban az ember fogalmát nemek szerint bontottuk nőre és férfira), majd ezt az eljárást folytatják addig, amíg olyan speciális fogalmakhoz jutnak amelyeknek vagy nincsenek további alárendeltjei, vagy az alárendelt speciális fogalmakat már nem kívánják alkalmazni. Ennek folytán növekvő “szélességű”, gúlaszerű, egymásnak alárendelt fogalmakat tartalmazó hierarchikus rendszer jön létre.

A mesterséges nyelven alapuló hierarchikus információkereső nyelvek **jelkészlete általában számjegyekből vagy elsősorban számjegyekből áll** (bár ez nem perdöntő – vannak példák a betűket alkalmazó rendszerekre is). Lényegesebb azonban, hogy **a szóképzés szabályai a fogalmi hierarchián** alapulnak. Ez azt jelenti, hogy ha egy fogalomnak egy – általában – számjegy(ek)ből álló “szó” (jelzet) felel meg, akkor az e fogalomnak alárendelt fogalmak “neve” (jelzete) ugyanezen számjegy(ek)ből és az ehhez alárendelt fogalmanként hozzárendelt további számjegy(ek)ből áll. A hierarchikus információkereső nyelvek szavaihoz az osztályozó rendszerek “szótáiraiban”, az ún. táblázatokban természetes nyelvű kifejezések járulnak, amelyek azonban csupán a – többnyire – számjegyekből álló “szó” természetes nyelvi megfelelőjét, magyarázatát jelentik. Az osztályozáshoz (vagyis az osztályozandó “valami” nevének lefordításához az információkereső nyelvre) elsősorban a jelzeteket a hierarchia rendjében tartalmazó részek (táblázatok) használatosak, de az információkereső nyelvi szavak természetes nyelvű megfelelőjéből általában (így az ETO esetében is) betűrendes (természetes nyelvű), lehetőség szerint szinonimákat is tartalmazó mutató is készül, amely utal a megfelelő, a hierarchiában könnyen lokalizálható jelzetre. A mutató azonban el is maradhat, egyszerűbb, kisebb fogalmi területre kiterjedő, hierarchikus

² Ez a kifejezés valójában minden információkereső nyelv szótárára alkalmas lehetne, annál is inkább, mert az osztályozás fő elvét, a fogalmi alá/fölérendeltséget – jóllehet csak utalásokkal – mindegyik kifejezni törekszik.

információkereső nyelvekhez a fentiek szerint mintegy magyarázatként járuló természetes nyelvi kifejezések léte esetén.

A hierarchikus információkereső nyelvek egyik fő problémája az, hogy **ugyanaz a fogalom különféle szempontok szerint lehetne felosztható** (vagyis a fogalomhoz különböző szempontok szerint lennének rendelhetőek annak specifikumait “létrehozó” ismérvek). Erre a hierarchikus információkereső nyelvek alkotóinak is számítaniuk kellett, és ezért – ha a gyakorlati információkeresés szempontjából vitatható módon is – egyes hierarchikus rendszerek esetében létre kellett hozniuk egy vagy több “másik hierarchiát”, a más felosztási (alárendelési) ismérveket képviselő, alternatív, az alaphierarchiát kiegészítő hierarchiákat. Ilyenek az ETO egészében érvényes földrajzi, nyelvi, népi, időbeli, műfaj szerinti és egyéni információkereső nyelvi “kifejezések”, vagyis ilyen az ún. általánosan közös alosztások önmagában is hierarchikus rendje. Emellett a hierarchia különböző ágaiban is kínál az ETO kiegészítő vagy alternatív fogalmakat jelölő információkereső nyelvi szavakat, nevezetesen a korlátozottan közös alosztásoknak az elsődleges hierarchiához csatlakozó saját hierarchiáját. Az 1. ábrán az általános és a kötőjeles korlátozottan közös alosztások alkalmazása látható.



1. ábra Általánosan és korlátozottan közös alosztások alkalmazása az ETO-ban. Az általánosan közös alosztást képviselő jelzetrészt aláhúzás, a korlátozottan közös alosztást jelölő jelzetrészt félkövér szedés jelöli.

*

Mind az általánosan közös, mind a korlátozottan közös alosztásoknak “betegsége”, hogy mivel az azokat kifejező információkereső nyelvi szavak az eredeti hierarchia szerinti, ún. “főtáblázati szavak” (jelzetek) végére kerülnek, a főtáblázat, tehát a hierarchikus osztályozó rendszer szerkezetét, hierarchiáját leképező, információkeresésre szolgáló eszközben (katalógusban) ezek az alosztások és így az általuk kifejezett fogalmak csaknem kereszteszóródnak. Gépi megoldások elvben lehetővé tennék e második vagy többedik hierarchiában képzett szavak (jelzetek) “keresési egyenjogúsítását”, de – egyelőre – sajnálatos módon csak igen ritkán lehet ilyen megoldásokkal találkozni.

Hasonlóképpen gyengéje e rendszereknek, hogy elvi alapjuk szerint minden fogalomnak több alárendelt fogalma van (lehet), és ugyanakkor minden fogalomnak csak egy fölrendelt fogalma, ami – különösen – összetett fogalmak esetén nem felel meg a valóságos összefüggéseknek. Ez a jelenség számos következtetésre vezet a valóságos osztályozási táblázatokban.

Az alosztások többé-kevésbé átgondolt rendszere által az értelmi összefüggések sokrétűbb kifejezésére kínálkozó lehetőségektől eltérően magában a “főtáblázatban”, tehát az egyértelműnek szánt rendszerben is előfordulnak csapongások, eltérő felosztási szempontok keveredései. Ezt szemlélteti a 2. ábra, amelyből kitűnik, hogy (és ez nem egyedülálló eset) ugyanazon a szinten több

felosztási szempont keveredik (a jogtudomány szakterületi felosztását – “Büntetőjog” – követi a tevékenység, azt pedig a résztvevők szerinti felosztás).

3.	Társadalomtudományok	Tudomány
34	Jog, Jogtudomány	Tudomány
343	Büntetőjog	Tudomány
343.1	Büntető eljárás	Tevékenység
343.16	A büntetőbíráskodásban résztvevő személyek	Személyek
343.19	A büntetőbírságok fajtái	Testületek

2. ábra. Különféle szempontok keveredése a felosztásokban az ETO példáján

*

Mielőtt elhagynánk a hierarchikus információkereső nyelvek területét, emlékezzünk meg néhány nyilvánvaló előnyükről. Ilyen előny az, hogy az alighanem legfontosabb értelmi összefüggést a bővebb-szűkebb (nem-faj) összefüggést hierarchiájukkal szemléltetően és követhetően kifejezik (ha egyoldalúan is). Elsősorban ennek következtében **kitűnően alkalmasak általánosabb, bővebb fogalmak szerinti tárgyköri információkeresésre**. További előny – legalábbis az ETO esetében – hogy a jelzetek különböző módon kombinálhatók, és ez által bizonyos “szövegösszefüggések” érvényesítésére alkalmasak. Végül előnyük, hogy e rendszerek, nyelvek nemzetközileg használhatók, mivel csak a fogalmak magyarázata és a mutató tekintetében kötöttek valamely természetes (nemzeti) nyelvhez.³

2.7 Mellérendelő információkereső nyelvek

A hierarchikus információkereső nyelvek és ezekre épülő katalógusok bonyolultsága és bizonyos előképzettséget feltételező használata, továbbá a specifikusabb, illetve nomenklatúrába foglalt kifejezések szerinti keresési igények folytán szükségszerű elvárásként alakult ki a **természetes nyelv használata információkereső nyelvként**. Ennek hívei általában arra hivatkoznak, hogy az információhasználó oldaláról e nyelvek nem igényelnek előzetes felkészülést, vagyis az ilyen nyelvekre épülő katalógusokban vagy gépi rendszerekben “mindenki tud keresni”. Azt azonban már láttuk a korábbiakban, hogy a természetes nyelv eredeti, “szűz” állapotában nem vagy kevéssé alkalmas információkeresés céljaira. Így egyrészt bizonyos megalkuvásokra, másrészt a természetes nyelvet “kordában tartó”, bizonyos mértékig szabályozó vagy legalábbis kiegészítő megoldásokra volt szükség erre a célra. Az ennek eredményeként létrejött információkereső nyelveket – megkülönböztetésül a hierarchikus nyelvektől – **mellérendelő nyelveknek**, a gyakorlatban pedig tárgyszavas nyelveknek nevezik.

Itt jegyezzük meg, hogy a következőkben tárgyalandó szemantikai információkereső nyelvek is a természetes nyelv szavainak a dokumentum, illetve a kérdés tartalma szerinti mellérendelésére épülnek. Mivel azonban fő sajátosságuk a fogalmak közti értelmi, jelentésbeli összefüggések rögzítése, megkülönböztetésül soroljuk őket az információkereső nyelvek külön csoportjába.

Elvben az eljárás igen egyszerű. A kereső (olvasó) kérdésének jellemzőit és a keresés tárgyául szolgáló dokumentumok, információk tartalmát (jellemzőit) egyaránt **a természetes nyelv egymás mellé rendelt szavaival (vagy kifejezéseivel) kell kifejezni**, és a kérdés így létrejövő keresőképét, illetve a “dokumentumképeket” kell **összehasonlítani**. Ez azonban a nyelvek fent már említett sajátosságai (egy kifejezés van több fogalomra, több fogalmat jelölő azonos kifejezés) miatt gyakorlatilag nem vagy csak nagyon nagy információvesztéssel működne. A természetes

³ Az ETO-t a magyar könyvtárakban széles körben alkalmazzák. Az alkalmazás lényeges problémája, hogy – amióta a rendszer gondozását nemzetközi konzorcium vette át az azóta megszűnt Nemzetközi Dokumentációs és Információs Szövetségtől (FID) – megszűnt a nemzetközi érvényű ETO-kiadványok publikálása. Az ETO középkiadásainak lényegében megfelelő, géppel olvasható hivatalos ETO-állomány magyar változatának elkészítése a fentiek írásakor még csak folyamatban van.

nyelvből valamilyen módon és valamilyen mértékben szabályozott (**kontrollált**) **nyelvet kell létrehozni**, és erre a nyelvre kell lefordítani a kérdés és a keresett dokumentumok (információk) tartalmát.

Mindez azt jelenti, hogy a “szabadon” megállapított, vagy – a dokumentumok esetében – a dokumentumból “kivont” szavak önmagukban csak jelöltek lehetnek arra, hogy helyet kapjanak abban a “tárgyszó-nyelvben”, amelynek szavait egymás mellé rendelve létre lehet hozni az összehasonlítandó “dokumentumképet” és a kérdés keresőképét. Ehhez természetesen össze kell állítani és szükség szerint karban kell tartani a tárgyszó-nyelv szótárát. Ennek első követelménye, hogy **ki kell küszöbölni a szinonimákat**, vagyis a szinonimapárokból vagy szinonimacsoportból ki kell választani a legcélszerűbbnek, leggyakrabban előfordulónak talált szót, kifejezést, a **kitüntetett tárgyszót**. A “kieső” szavakról utalni kell kitüntetett párjukra (például eb lásd kutya), illetve összetett kifejezések esetén a második taggal kezdődő formáról az első taggal kezdődő formára (például személyi számítógép lásd számítógép, személyi). Szemantikai információkereső nyelvekben az utóbbi fajta utalás szükségtelen és nem is szabad alkalmazni, mert meg kell adni a számítógép és a személyi számítógép közötti generikus kapcsolatot, miáltal a számítógép alatt alárendeltként megtalálható a személyi számítógép.

Nyilván felmerül a kérdés, hogy miért nincs szükség vagy miért van kevésbé szükség ilyen utalásokra a hierarchikus információkereső nyelveknél, mint a mellérendelőknél. Az ok – bár a hierarchikus rendszerek mutatóiban, sőt, adott esetben táblázataikban sem nélkülözhetők az utalók – könnyen érthető: a hierarchikus rendszerekben az információkereső nyelvi szó elhelyezkedése a hierarchiában viszonylag könnyen megtalálhatóvá teszi a “kitüntetett” szót.

A szinonimák kiküszöbölése, illetve a homonimák értelmezése önmagában még nem elegendő. Megemlékeztünk róla, hogy az **értelmi/tartalmi összefüggések** mellőzése információveszteségre vezet. A hierarchikus rendszerek esetében magából a hierarchiából – ha nem is differenciáltan – következik a két legfontosabb értelmi összefüggés: általában a nem és a faj, illetve ritkábban (a két összefüggés-fajta megkülönböztetése nélkül) az egész és a rész összefüggés. A mellérendelő információkereső nyelvekben ennek megfelelően használható eszköz nem áll rendelkezésre. Helyette – nem differenciálva az értelmi összefüggés jellege és iránya (például bővebb/szűkebb vagy egész/rész) szerint – a mellérendelő nyelvekben a **“lásd még” típusú utaló** (például tál lásd még edény, illetve edény lásd még tál vagy tányér vagy fazék vagy lábas vagy üst) oldja meg – ha nem is az információkeresés minden igényét kielégítően – az értelmi összefüggések nyomon követését.

A mellérendelő információkereső nyelveknek nevezett tárgyszórendszerek minőségét befolyásolhatja az is, hogy a tárgyszóként választott szavak milyen mértékben összetettek. Az összetett szavak – hacsak második részükről nincs utalás az első résszel kezdődő formára – nehezebben megtalálhatók, de ugyanakkor tartalmilag, fogalmilag lényegesen kifejezőbbek.

Itt említjük még meg, hogy igen gyakran – különösen a számítógépes információkeresés adventje óta – előfordul természetes nyelvi szavak, úgynevezett “szabad tárgyszavak” (valójában nem tárgyszavak, hanem szabad szavak vagy szabad szövegszavak) alkalmazása is, a fentiekben említett mindennemű utalás nélkül, vagyis tárgyszórendszer nélkül, esetleg egy vagy több információkereső nyelv használatával történt osztályozás/indexelés és ennek megfelelően végzett keresés kiegészítőjeként. Kizárólagos alkalmazásuk – bár az eredmények egyes megoldásokkal javíthatók – kedvezőtlenül hat az információkeresés eredményességének mutatóira, mindenekelőtt a keresési veszteségre, ugyanakkor, amikor egyéb esetekben a zajt is növelheti. Mindazonáltal egyszerűsége miatt a “szabad tárgyszavak” használata meglehetősen elterjedt és – különösen kiegészítő megoldásként – nem is haszontalan.

2.8 Szemantikai (relációs) információkereső nyelvek

A mellérendelő információkereső nyelvek vázlatos áttekintéséből látszik, hogy minél inkább kiegészítik alkalmazásukat különféle, a tárgyszavak közötti értelmi összefüggéseket kifejező utalók alkalmazásával – annál eredményesebbé válhat az e nyelvek alkalmazásával végzett tárgyköri

jellemzés és információkeresés. A gyakorlatban ez a felismerés vezetett el a természetes nyelv szavainak, kifejezésének használata tekintetében a tárgyszavas rendszerekhez hasonló, de az értelmi összefüggéseket széleskörűen és differenciáltan kibontó, a zajt és a veszteséget – megfelelő alkalmazásuk esetén – minimalizáló **szemantikai vagy más néven relációs információkereső nyelvek** és szótáraik kialakulásához. Mind a szemantikai, mind a relációs jelző azt fejezi ki, hogy ezek a nyelvek a fogalmak (és az azokat kifejező természetes nyelvű kifejezések) értelmi, (jelentésbeli, vagyis szemantikai és morfológiai) összefüggéseinek, vagyis relációinak minél teljesebb kifejezésére törekednek. E törekvés funkcionálisan hasonló a tárgyszórendszerek “lásd még” és “lásd” utalóival elérni kívánt eredményekhez, de azoknál hasonlíthatatlanul finomabb, pontosabb eszközt eredményez.

A szemantikai információkereső nyelvek lexikai egységeit (“tárgyszavait”), vagyis a valamely fogalom leírására, képviselőre kiválasztott szavakat vagy kifejezéseket **deszkriptoroknak**, szótáraikat pedig – jelezvén, hogy ezek a fogalmak “kincstárai” – **tezauruszoknak**⁴ nevezzük. (A szavaik alapján nevezik még ezeket a nyelveket deszkriptoros nyelveknek is, ahogy a tárgyszavak alapján tárgyszavas nyelvekről is szoktak beszélni.)

Mint említettük, a szemantikai információkereső nyelvek sokrétűen kifejezik a természetes nyelv kifejezéseivel megjelölt fogalmak kapcsolatait, értelmi összefüggéseit. Ehhez természetesen a tezauruszok megalkotóinak számos dologban dönteniük kell. Az egyik – ha úgy tetszik az első – ilyen döntés szinonimák esetében merül fel (akárcsak a tárgyszavas rendszerek “lásd” utalóinak megalkotásakor). A deszkriptorokkal szinonimaviszonyban (relációban) lévő, tehát azonos jelentésű, vagy legalábbis a deszkriptor jelentésével kielégítően jellemezhető rokonjelentésű kifejezések a nemdeszkriptorok, amelyeket a deszkriptorokkal kölcsönös utalás kapcsol össze, egyrészt jelezve, hogy a nemdeszkriptor helyett melyik deszkriptor használandó, másrészt, hogy a deszkriptort milyen nemdeszkriptor(ok) helyett kell használni. A szinonimák tekintetében előfordul, hogy nem egy másik kifejezés önmagában szinonima, hanem két kifejezés együttese, logikai szorzata (például védőoltás – lásd immunizálás és oltás). Hasonlóképpen előfordul, hogy egy nemdeszkriptornak minősített kifejezésről más kifejezések vagylagos együttesére (logikai összegére) utalnak (például felsőoktatási intézmény lásd egyetem vagy főiskola).

E nyelvek és szótáraik legfontosabb vonása azonban az, hogy teljességre törekedve és lehető legnagyobb differenciáltsággal kimutatják tartalmuk elemeinek (szókincsüknek) az egyéb információkereső nyelvek által is kimutatott szinonima-reláción kívüli értelmi, jelentésbeli összefüggéseit.

Fő részük betűrendbe sorolt úgynevezett **deszkriptorcikkek**ből és **nem-deszkriptorcikkek**ből **épül fel**. Már említettük, hogy a deszkriptor két azonos vagy közeli rokon értelmű szó vagy kifejezés (a továbbiakban: kifejezés) közül az, amelyiknek használata mellett döntöttünk, és a nemdeszkriptor e kifejezéspárból az, amelynek használatát mellőzzük.

A **deszkriptorcikk tartalmazza** magát a deszkriptort (mint látni fogjuk, más deszkriptorokat is tartalmaz, ezért azt a deszkriptort, amelynek deszkriptorcikkéről “szó van”, amelynek összefüggéseit (relációit) a deszkriptorcikkben található nemdeszkriptor(ok)kal és más deszkriptorokkal a deszkriptorcikk kimutatja, vezérdeszkriptornak nevezzük. Emellett tartalmazza a deszkriptorral különféle fajtájú értelmi, jelentésbeli, szemantikai összefüggésben álló deszkriptorokat, az összefüggés jellegét kifejező szabványos rövidítésekkel. Mindezekhez szükség esetén a vezérdeszkriptor természetes nyelvű definíciója vagy használati utasítása, alkalmazásának története, forrása stb. csatlakozhat, továbbá – olyan tezauruszok esetében, amelyek a deszkriptorokat ún. szakcsoportokba sorolják, a megfelelő szakcsoport megjelölése járulhat.

A **nemdeszkriptorcikk** a nemdeszkriptort és a helyette használandó deszkriptort, vagy a megadott összefüggés szerint (vagy, illetve és) alkalmazandó deszkriptorokat tartalmazza.

A vezérdeszkriptorral, illetve (nemdeszkriptorcikk esetén) a nemdeszkriptorral értelmi összefüggésben álló nemdeszkriptorok és deszkriptorok relációinak megjelölésére a magyar szabvány szerint az alábbiakban felsorolt megjelöléseket alkalmazzák a deszkriptorcikkekben,

⁴ Más írásmód szerint tezauruszoknak

illetve nemdeszkriptorcikkekben (a felsorolásban a vezérszkriptorról és a megjelölt deszkriptorokról szólva természetesen a vezérszkriptor, illetve a deszkriptor jelentéséről van szó). A vezérszkriptorhoz, illetve a nemdeszkriptorhoz különböző típusú megjegyzések és magyarázatok kapcsolódhatnak (az alábbiakban csak az általános magyarázatot és a használati megjegyzést tüntettük fel).

Deszkriptorcikkekben:

- M:** – a vezérszkriptorra vonatkozó általános magyarázat, meghatározás
- H:** – a vezérszkriptorra vonatkozó használati megjegyzés
- H** – a vezérszkriptort kell használni a megjelölt nemdeszkriptor helyett
- H&** – a vezérszkriptort kell használni a megjelölt nemdeszkriptorok együttes használata helyett
- HV** – a vezérszkriptort kell használni a megjelölt nemdeszkriptorok bármelyike helyett
- A** – a megjelölt deszkriptor alárendeltje (faja) a vezérszkriptornak
- F** – a megjelölt deszkriptor fölrendeltje (neme) a vezérszkriptornak
- T** – (totum) a megjelölt deszkriptor egésze a (részét képező) vezérszkriptornak
- P** – (pars) a megjelölt deszkriptor része az (egészét képező) vezérszkriptornak
- R** – a megjelölt deszkriptor rendeltetése, oka, eredete a vezérszkriptornak
- E** – a megjelölt deszkriptor eszköze, okozata, eredménye a vezérszkriptornak
- X** – a megjelölt deszkriptor közelebbről meg nem jelölt értelmi összefüggésben áll vezérszkriptorral
- ≠** – a megjelölt deszkriptor azonos írásképpű, mint a vezérszkriptor, de más értelmű

Nemdeszkriptorcikkekben:

- M:** – a nemdeszkriptorra vonatkozó általános magyarázat, meghatározás
- H:** – a nemdeszkriptorra vonatkozó használati megjegyzés
- L** – a megjelölt deszkriptort kell használni a nemdeszkriptor helyett
- L&** – a megjelölt deszkriptorok együttesét (illetve közülük többet) kell használni a nemdeszkriptor helyett
- LV** – a megjelölt deszkriptorok valamelyikét kell használni a nemdeszkriptor helyett

Deszkriptorcikk és nemdeszkriptorcikk példája a 3. ábrán látható. Összehasonlításként az ábra bal oldalán két lehetséges tárgyszócikk látható.

*

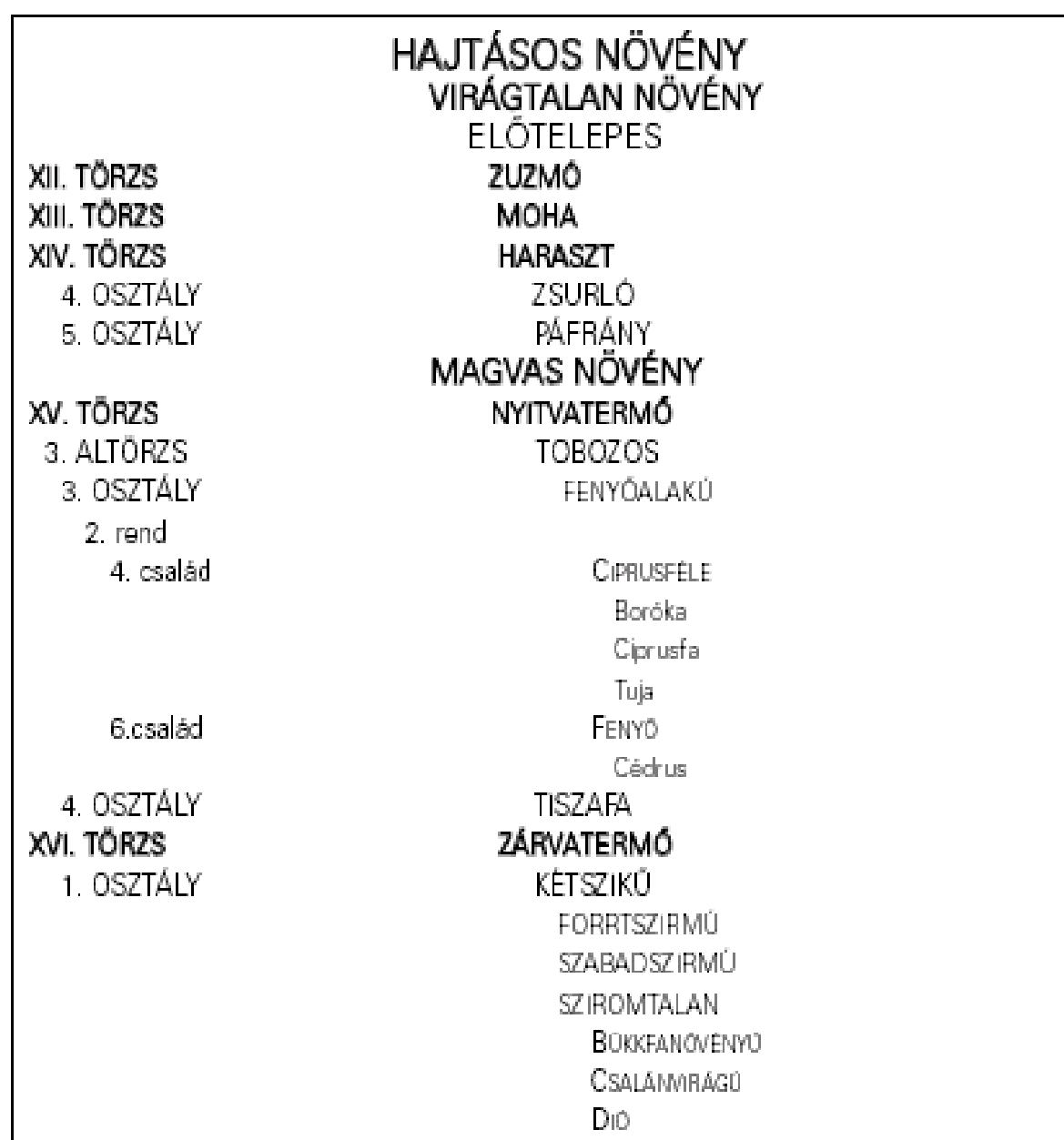
államigazgatás	államigazgatás
M: Az államhatalom felsőfokú rendelkező, végrehajtó, szervező, irányító tevékenysége	Lásd még államjog vagy felügyelet vagy hatóság vagy közigazgatás
H: Az „-igazgatás” és „Államigazgatás” összetételek egy része a „Közigazgatás-” kezdetű lexikai egységeknél található	
H felsőfokú igazgatás	
HV igazgatás	
F irányító tevékenység	
A felügyelet	
hadügy	
közigazgatás	
külügy	
rendészet	
T államjog	
civilizáció	
P ügyvitel	
R közjó	
E hatóság	
X államigazgatási eljárás	
felsőfokú igazgatás	felsőfokú igazgatás
L államigazgatás	Lásd államigazgatás

3. ábra

Szabványos teauruszcikk (bal oldalon) példája az OSZK-teaurusz/Köztauruszból. Alul a nemdeszkriptorcikk, illetve a “lásd utalás” példája látható. A jobb oldalon ugyanannak a két fogalomnak a lehetséges tárgyszócikke látható (átlagos tárgyszórendszert véve alapul).

*

Megemlítendő még, hogy a teauruszok általában több részből állnak. Fő részük a deszkriptorcikkeket és nemdeszkriptorcikkeket betűrendben tartalmazza. Vannak – lehetnek azonban egyéb részeik is. Elsősorban a Magyarországon készült információkereső teauruszok jelentős részének sajátossága, hogy különálló rész(ek)ben, úgynevezett taxaurusz(ok)ban tartalmazzák a teaurusz tárgykörébe tartozó egy vagy több tudományág által megnevezési rendszerbe (taxonómiába) foglalt megnevezéseket (deszkriptorokat). Mivel ezek a megnevezési rendszerek általában hierarchikusak, a taxaurusz betűrendes részéhez a deszkriptorok hierarchiáját közvetlenül (explicite) is tartalmazó/ábrázoló mutató is készül. Ilyen – elvben készülhet a teauruszhoz is, de a teauruszokon belüli hierarchikus kapcsolatok polihierarchikus jellege miatt nagy nehézségekkel (tehát nem csak a kifejezetten hierarchikus részhez) (a taxaurusz részlete látható a 4. ábrán).



4. ábra Az OSZK-tezaurusz/Köztaurusz taxauruszának részlete

A tezauruszok a hierarchikus rész mellett tartalmazhatnak más részeket is: például a **deszkriptorokat tárgyköri csoportokba soroló szakcsoportos részt (mutatót)**, amelynek fő hasznosítása olyan deszkriptorok, illetve nemdeszkriptorok keresése, amelyekhez, illetve amelyekhez közelálló, de betűrendben nem talált lexikai egységekhez a szűkebb szakterület kifejezéseinek áttekintése útján lehet eljutni. Lehetnek a tezauruszoknak a deszkriptorok (és esetleg nemdeszkriptorok) összefüggéseinek rendszerét hálószerűen, gráfokkal bemutató **grafikus részei** is. Ez utóbbiak gyakran a tezaurusz szerkesztése során készülnek.

2.9 Az információkereső nyelv és az információkereső rendszerek fejlődő kapcsolata

Az információkereső nyelvről eddig írottakban elvontan volt szó a kérdés és a keresett információ “információkereső nyelvre fordításáról”, illetve a kérdés és a keresett információt rejtő halmaz leírásainak összehasonlításáról. A valóságban ez az összehasonlítás évszázadokig csak úgy volt lehetséges, hogy a kereshetővé tett információkat vagy – utóbb – ezek leírását meghatározott szempontok (például szerzők neve vagy dokumentumok címe vagy éppenséggel valamely – igen általános, “sekély” tárgyköri információkereső nyelv) szerint fizikailag elrendezték, majd a kérdés hasonló ismérvfajták szerint, hasonló “nyelven” történt megfogalmazása után **megkeresték azt a fizikai helyet**, ahol a keresett információ található volt. Ez a helyzet alapjában nem változott meg még a cédulakatalógusokkal sem, csupán az elrendezési ismérvek köre (vagyis az alkalmazott információkereső nyelv) vált differenciáltabbá. Ugyanakkor egyre több problémát okozott, hogy a leírásokat általában csak **egyetlen szempont szerint** lehetett egy, esetleg több “helyre” (de e helyek közötti kapcsolat nélkül vagy annak problematikus nyomon követését biztosító kezdetleges eszközökkel) elhelyezni, tehát összetettebb kérdéseket meg lehetett ugyan információkereső nyelven fogalmazni, de az elrendezett információgyűjteményben (katalógusban) “vagy itt, vagy ott” lehetett keresni.

Az első olyan technikai eszközök, amelyek lehetővé tették az információkeresési ismervek szerinti együttes, koordinált információkeresést, a Taube-féle, ismervenként elrendezett dokumentumazonosítókat tartalmazó **“Taube-kártyák”**, másnéven **uniterm kártyák** voltak, majd ezeket követték a **kézi lyukkártyák**, peremlyukkártyák, réslyukkártyák és – az ismervek szerinti invertált fájlok prototípusaként – a **fénylyukkártyák**. Rövid gépi lyukkártyás és mutatók előállításához alkalmazott lyukszalagos intermezzo után a **számítógépes információkeresés nyitotta meg az utat** az információkereső nyelven leírt ismervek szerinti, teljes információállományban végzett keresés előtt. E kezdetek kiteljesedéséhez a számítástechnika korszerűsödésére, a személyi számítógépek elterjedésére, majd – különösen, de egyben új problémákat felvetően – az internet létrejöttére kellett “várni”. Ezzel párhuzamosan az is egyre inkább elterjedő gyakorlattá vált (bár ez a “manuális korban” sem volt ismeretlen; lásd egyebek között “katalogizálás a kiadványban” – cataloguing in publication – CIP), hogy a hagyományos, illetve az elektronikus, köztük a hálózati lelőhelyű **dokumentumokat létrejöttükor ellátják információkereső nyelven megfogalmazott leírásukkal**. Sajnálatos és jellegzetes módon azonban az e célra alkalmazott információkereső nyelvek gyakran primitívek, ami a kereséskor téves eredményekre (“zajra”), illetve információvesztésekre vezet.

Nehezen prognosztizálható, hogy a fejlődés milyen irányokba vezet. Ami az online keresési technikákat illeti, nem egykönnyen képzelhető el a jelenleginél lényegesen kedvezőbb technikai környezet az információkereső nyelvek eredményes alkalmazásához. Több mint valószínű azonban, hogy erős fejlődésnek indul az automatizált indexelés, osztályozás, vagyis információkereső nyelvekre fordítás – remélhetőleg lényeges minőségi engedmények nélkül.

3. Az információkeresés technikája

3.1 A keresés folyamata

A könyvtári információkeresés végezhető kézzel (manuálisan) cédulakatalógusban, dokumentumokban vagy akár másokat megkérdezve. Végezhető mechanikai eszközökkel és készülékekkel (lyuk- és fénylyukkártyák, mikrofilmes berendezések) és számítógépek segítségével. A mechanikai eszközökkel és készülékekkel, továbbá az olyan programrendszerekkel végzett keresést, amely nem teszi lehetővé a számítógéppel folytatott kommunikációt, offline keresésnek nevezik, és a könyvtári információkeresésben alig van már szerepe.

Online információkereső rendszerekben a felhasználó közvetlen – interaktív – kapcsolatot teremthet a számítógép segítségével a dokumentumleírások (olykor a digitális vagy digitalizált dokumentumok) állományával.

Ma az információkeresésnek szinte mindig van online szakasza. Ezért az információkeresés technikai kérdéseit ebben a könyvben az online információkeresés keretében tárgyaljuk – megjegyezve, hogy nem minden itt tárgyalt művelet és fogalom kizárólag az online keresés, hanem gyakran általában az információkeresés sajátossága.

A dokumentumleírásokat (nevezik ezeket dokumentációs egységeknek, dokumentumrekordoknak, dokumentumtételeknek, a továbbiakban az utóbbit fogjuk használni) adatbázisokba szervezve tárolják; a könyvtári információkeresés hagyományosan ezekben játszódik le.

Vannak olyan online keresőrendszerek, melyek segítségével magukat az eredeti dokumentumokat lehet megkapni. Ezeket a dokumentumokat vagy a nyomtatott dokumentumok digitalizálásával állítják elő, és digitális formában tárolják, vagy eleve digitálisan készülnek. A digitálisan rögzített dokumentumok ma elsősorban az interneten keresztül érhetők el.

Az online információkereső rendszerben a számítógép és a felhasználó között kétirányú a kommunikáció, mely a beviteli, illetve kiviteli berendezéseken (billentyűzet, egér, illetve képernyő, nyomtató) és a kommunikációs csatornán keresztül valósul meg (mint amilyen a telefonvonal és a távközlési eszközök segítségével fenntartott távolsági adatátviteli vonalak), és a kommunikáció fázisai (így a kérdések és a válaszok) a számítógéphez kötött képernyőn jelennek meg.

Az interaktív hozzáférés közel hozza a végfelhasználót az adatbázishoz vagy az adatok egyéb állományához. Ettől azonban csak látszólag egyszerűbb a dolga. Ha ugyanis nem információs

szakember, kevésbé járatos a mindenkori információkereső rendszer eszközeinek használatában, akkor keresési eredményei ennek megfelelően tökéletlenek lesznek.

Az információkeresés **legfontosabb művelete az összehasonlítás**. Egész általánosan megfogalmazva: a felhasználó információkereső nyelvre lefordított kérdését össze kell hasonlítani a tárban található, ugyancsak információkereső nyelven megfogalmazott információkkal. Azok a tárolt információk, melyek eme összehasonlítás eredményeként megegyeznek a kérdés információjával, vezetnek el a találatokhoz. Online rendszerekben az összehasonlítást a program végzi. **Összehasonlításkor a kérdés és a dokumentum ismérveinek összevetésére kerül sor** (a következő fejezetekben ezt a műveletet tovább pontosítjuk a keresőprofil fogalmának bevezetésével). (Az összehasonlítással, az információkereső nyelvre való fordítással és az ismérvekkel részletesen a "2. Az információkereső nyelv és szótára" című fejezetben részletesen foglalkoztunk).

A fordítás sikere azon is múlik, milyen a párbeszéd a rendszerrel, mennyi segítséget kap a rendszertől a felhasználó. E téren alig vannak különbségek a könyvtári rendszerek között; a laikus végfelhasználónak nyújtott segítségek (súgók, "helpek") nagyon szegényesek. Az interneten működő keresőszolgáltatásokban már több segítség található, de ezek a tájékoztatások sokszor nincsenek eléggé tekintettel a felhasználóra, a keresési példák, melyekre a leginkább szükség lenne, ezekben a segítségekben is jóformán teljesen hiányoznak.

Az se mindegy, hogy **milyen információkereső nyelvet használnak** az adott rendszerben, és milyen minőségű a dokumentumok feldolgozása.

Miután megtörtént az összehasonlítás, és a rendszer megjelenítette a találatokat, ezt a "választ" a felhasználónak meg kell értenie, ami azt jelenti, hogy le kell fordítania a saját gondolataira (ami szintén "fordítási" művelet, noha ugyanúgy nem tudatos, akár a fordítottja, és legalább annyi félreértés forrása, mint az előző fordítások bármelyike).

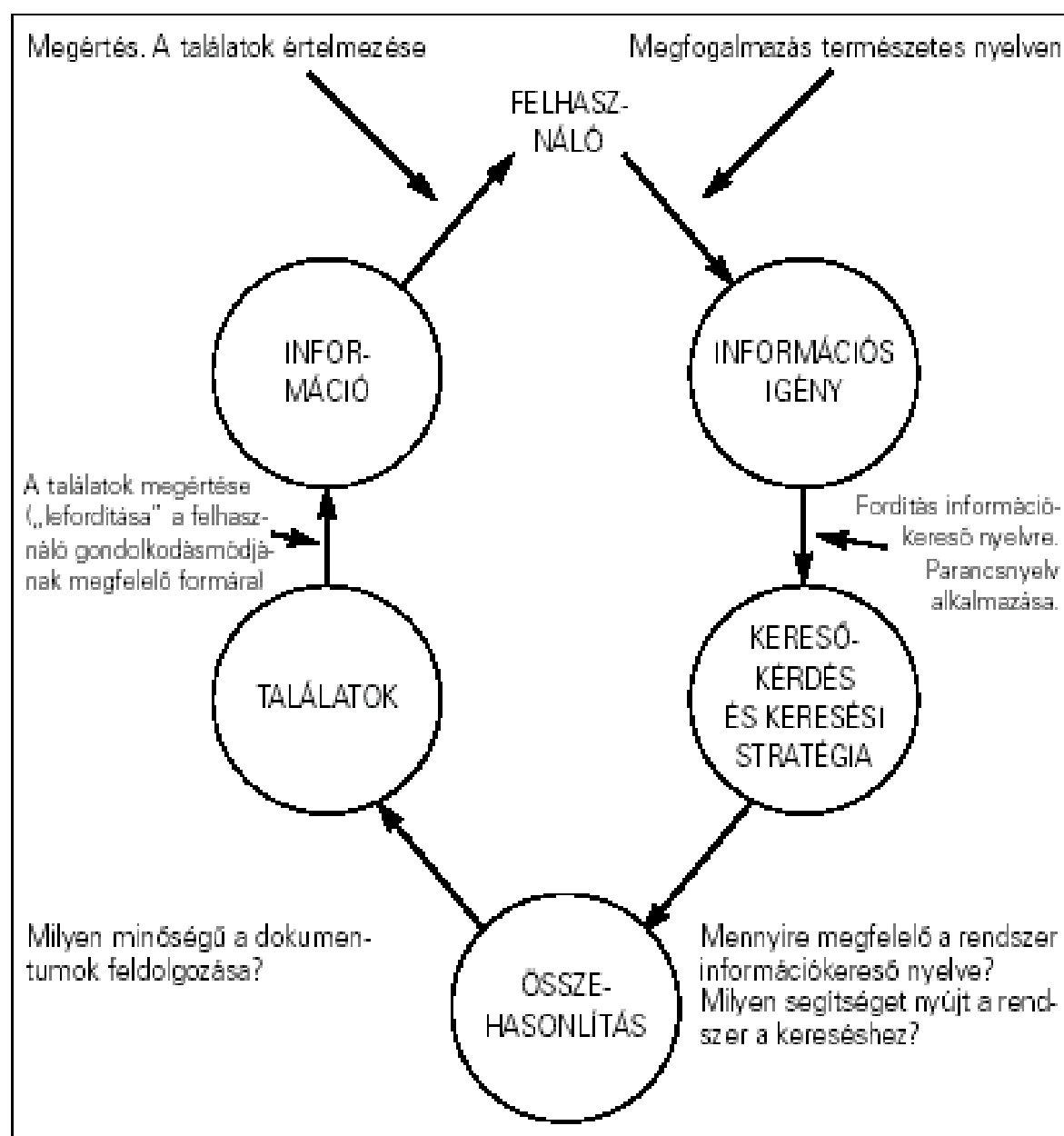
A felhasználónak végül értelmeznie is kell a válaszokat képviselő találatokat, a kapott dokumentumtöredékeket is. Ennek érdekében jó, ha ismeri a kereséshez használt rendszert gyengeségeit, előnyeit, mert ezek befolyásolják, hogy mit kap találatként.

Ismerni kell továbbá azokat a parancsokat (billentyűzetről vagy egér segítségével megadott utasítások a programrendszer számára), melyeket az információkereső nyelvre lefordított keresőkérdés bevitelekor használni kell. **E parancsok összessége, a parancsnyelv** is különböző lehet rendszerenként (részletesen lásd a "3.5 A parancsnyelv" című fejezetben).

Mindebből látható, hogy az online információkeresés önmagában nem oldja meg a kereséskor felmerülő lényeges nehézségeket. Amit önmagában megold, az csak a keresés kényelme és gyorsasága.

A teljes folyamatot az 5. ábrán szemléltetjük.

*



5. ábra. Az információkeresés teljes folyamata

*

3.2 Könyvtári adatbázisok rekordszerkezete és mutatói

3.2.1 A rekordszerkezet

Az online keresés a könyvtári és más rendszerekben mindig meghatározott logikai rekordszerkezetű dokumentumtégeket eredményez. Ezek – szemben az eredeti dokumentumokban szereplő elsődleges információkkal – másodlagos információk (nevezik hivatkozási információknak is). (A weben kereshető HTML-dokumentumok szerkezetével az “5. Keresés az interneten” című fejezetben foglalkozunk).

Számítástechnikai szempontból a dokumentumrekord kifejezést használják. (Nem szerencsés pusztán “rekordról” vagy “tételről” beszélni, mert könyvtári rendszerekben nem csak dokumentumrekordok/tételek fordulnak elő. A 6. ábrán például a “Móricz Zsigmond (1879–1942)” szintén rekord vagy tétel: egységesített besorolási adattétel.)

Könyvtári nézőpontból a dokumentumtételek dokumentumleírások. Az a részük, amely a bibliográfiai leírás szabványainak követelményei szerint készül, a bibliográfiai leírás. Ha a bibliográfiai leírás kiegészül különféle egyéb katalogizálási adatokkal (például besorolási adatokkal, raktári jelzetekkel), akkor katalógustételről beszélünk.

Az alábbi 6. ábrán az Országos Széchényi Könyvtár adatbázisának egyik megjelenített dokumentumtége látható; bal oldalt szabványos formában (eltekintve az alábbiakban még megmagyarázott aláhúzásoktól), jobb oldalt pedig HUMARC adatcsere-formátumban.

*

Bibliográfiai:	HUNMARC:
Márcia Zsigmond (1879-1942)	000 - \$aMÁRCIA ZSIGMOND
Leírás: (tétel)	001 - 7395
ISBN 963-571-418-1	005 - 200010050902055
ISBN 963-571-418-1	008 - 920627a1991 ***\$a\$u\$r***j*****0000i*pa1*1
ISBN 963-571-418-1	010 - ** \$a963-571-418-1 \$jlibrett Scar nélkül
ISBN 963-571-418-1	040 - ** \$aMNB HUN \$cMNB HUN
ISBN 963-571-418-1	041 1* \$apapal \$bun
ISBN 963-571-418-1	072 *1 \$aA
ISBN 963-571-418-1	080 ** \$a087.5(884.1)
ISBN 963-571-418-1	100 10 \$aMárcia \$jZsigmond \$d1879-1942
ISBN 963-571-418-1	240 10 \$aLeírás (tétel) \$ileggyel
ISBN 963-571-418-1	245 10 \$aMárcia \$jZsigmond \$c[aut. Zsigmond Márcia] \$j[prezsk. Jerzy Handzlik] \$d[il. Emese Bárczi]
ISBN 963-571-418-1	260 1* \$a[Budapest] \$b[Livroservo] \$c[1991] \$d[Budapest] \$e[All. Ny.]
ISBN 963-571-418-1	300 ** \$alapozó (8 p.) \$bill., színes \$c10x14 cm
ISBN 963-571-418-1	300 ** \$aVerses képeskönyv
ISBN 963-571-418-1	700 10 \$aBárczi \$jEmese
ISBN 963-571-418-1	11 \$aHandzlik, \$jJerzy \$cford.

6. ábra.

Dokumentumtétel megjelenítése és adatsere-formátuma.

Webhelye: <http://w3.oszk.hu/hektor.htm>

Az ábra bal oldalának felső sorában a szerző neve látható egységesített besorolási alakban, amit a cím követ. A tétel alján az illusztrátor és a fordító neve látható adathiányos egységesített besorolási adatként.

A tétel alján a kiadvány nemzetközi könyvazonosító-száma (ISBN) és az ETO-jelzet látható. A kiadónévek (itt a “Livroservo”, helyesen “Libroservo” – manapság nem ritka beviteli hiba, amire célszerű kereséskor felkészülni) alapján is keresni lehet, de ez nem egységesített besorolási adat.

A jobb oldalon a HUNMARC adatsere-formátum szerinti mezők hívójelei, majd a meghatározott kezelést, értelmezést jelző indikátorértékek, közvetlenül az adat értéke előtt pedig a megfelelő almező(k) azonosítói láthatók (az utóbbiakat a \$ jel előzi meg. **A mezők hívójelei, az indikátorok és az almezők azonosítói a formátum tartalomjelölői.**

Az egységesített besorolási adatok és az ETO-jelzet nélküli (de az ISBN-t tartalmazó) rész a dokumentumtételen belül a szűken vett bibliográfiai leírás. Csak az aláhúzott almezők tartalma alapján lehet keresni (például az 500**\$a megjegyzésmező “Verses képeskönyv” tartalma alapján általában nem lehet), de az egyes mezők kereshetősége mindig a kezelőrendszer függvénye.

*

3.2.2 Az online mutató

A példaként választott dokumentumleíráson belül az aláhúzások jelölik azokat az adatelemeket, amelyek szerint a tétel kereshető. Ezek tehát a tétel **ismérvei**.

A kezelőrendszerrel függően szövegen belüli szavak (tehát nem aláhúzott adatelemek) szerint is lehet keresni. Ebben az esetben ezek a kereshető szavak is ismérvek.

A mai online adatbázisokban azért lehet gyorsan, hatékonyan keresni, mert a dokumentumtégeket és a dokumentumtétel ismérveit (például főcímet, személynevet, ETO-jelzetet, deszkriptorokat) célszerűen egymáshoz rendezve tárolják.

Az ismérvek és dokumentumtétel egymáshoz rendelésének logikai formáját nevezzük **dokumentum-ismérv mátrixnak** (7. ábra). Bal szélső függőleges oszlopában a dokumentumtétel elsődleges azonosítói szerepelnek (a 6. ábra dokumentumtételében ilyen a 001 hívójelű adatelem 7395 értéke). A vízszintes sorokban az előforduló ismérvek találhatóak (a 6. ábra

bibliográfiai rekordjában ilyenek az aláhúzott adatelemeknek megfelelő 020a, 100, 260b, 240 és 700 hívójelű adatelemek).

A 6. ábrán látható dokumentumtételnek valójában az aláhúzottnál több ismérve van, mivel a példának választott rendszerben a 008 mező 7–10 pozícióján szereplő megjelenési év (1991) és a 15–16. pozícióján szereplő kiadó ország (hu) is kereshető adatelemek. Ezek az ismérvek azonban a bal oldali dokumentumleírásban nem jelennek meg.

*

	Ismérvek					
	I_1	I_2	I_3	I_j	...	I_n
D_1	a_{11}	a_{12}		a_{1j}		a_{1n}
D_2		a_{21}	a_{23}	a_{2j}		
...				...		
D_i	a_{i1}	a_{i2}	-	a_{ij}		a_{in}
...						...
D_m	a_{m1}	a_{m2}	a_{m3}	a_{mj}		a_{mn}

Az a elem értéke 1, ha a hozzá tartozó I ismérvet használják a rekordban, és 0, ha nem használják.

7. ábra. Dokumentum-ismérv mátrix

*

A vízszintes felosztás olyan kereshető tételeket jelent, melyek egy-egy dokumentumra vonatkoznak, és az ezzel a dokumentummal összekapcsolt ismérveket tartalmazzák. A függőleges bontás viszont olyan tételeket jelent, amelyek egy-egy ismérvet tartalmaznak, valamint azokat a dokumentumtégeket, amelyekhez az adott ismérvet hozzárendelték. A mátrix ezen kívül még többféle módon is fölbontható, s ezek alapján belőle ötféle tétel alakítható ki.

- (1) dokumentum-hozzáférésű tétel, például $D_1: I_1, I_2, I_j$
- (2) ismérv-hozzáférésű tétel, például $I_3: D_2, D_m$
- (3) csoportos ismérv-hozzáférésű tétel, például $I_1 + I_3: D_1, D_2, D_i$
- (4) csoportos dokumentum-hozzáférésű tétel, például $D_1 + D_2: I_1, I_2, I_3$
- (5) elemi tétel: $I_3: D_2$

Az első típus (1) a **dokumentum-hozzáférésű tétel**. Ezek összessége alkotja logikailag a dokumentumtétel azonosítói szerint rendezett állományát (nevezik “bibliográfiai törzsfájlnak” is). Tartalmazza a dokumentumok teljes tételét.

A korszerű relációs adatbázis-kezelő rendszerek által kezelt adatbázisokban az ismérvek (egységesített besorolási adatok, ETO-jelzetek, tárgyszavak, deskriptorok) nem részei a bibliográfiai törzsfájlnak, hanem önálló fájlokat alkotnak, melyeket az adatbázis-kezelő rendszer kapcsol össze.

A második típus (2) az **ismérv-hozzáférésű tétel**. Ezek összessége alkotja az ún. **invertált (alternatív) indexet vagy állományt**. Invertálnak azért nevezik, mert a dokumentumok rendezettségéhez képest egy másik (“fordított”, ui. **ismérvek szerinti) rendezési alternatívát képvisel**.

Ezek az invertált indexek a teljes adatbázishoz képest sokkal kisebb állományok. Bennük az ismérvek értékei (a 9. ábrán látható ilyen invertált index (mutató) a megfelelő dokumentumrekord azonosítójával vannak összekapcsolva.

A rendszer először megkeresi az ismérvet az indexben, és az ismérvhez kapcsolt mutató segítségével jut el a dokumentumrekord(ok)hoz a főállományban. (Mindez lényegesen gyorsabban játszódik le, mintha a dokumentumokat kellene sorra végignézni, hogy tartalmazzák-e a kereséshez használt ismérvet, azaz dokumentum-hozzáférésű tételek alapján szerveznék meg az adatbázist.) Az

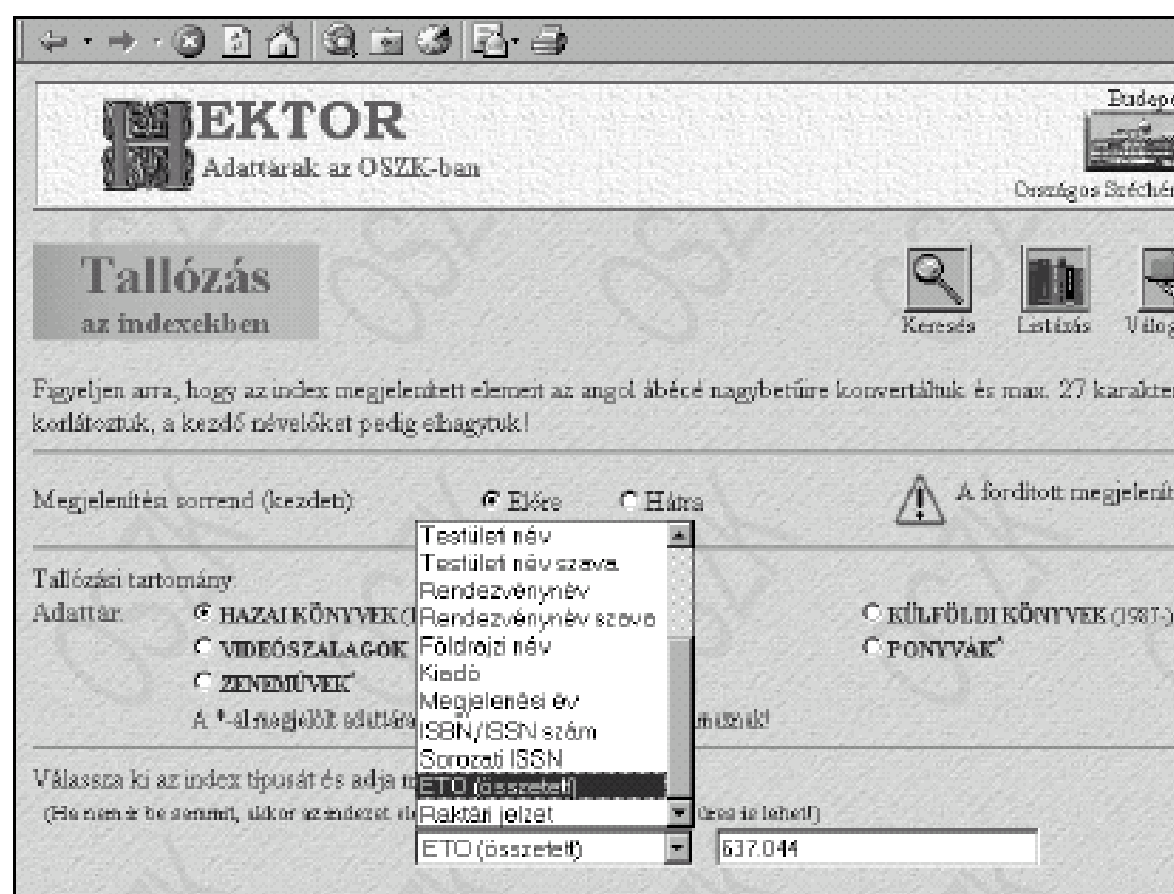
elérésnek az itt leírt módját nevezik közvetlen, véletlen vagy tetszőleges elérésnek, és az ilyen invertált indexeket számítástechnikai nézőpontból nevezik indexszekvenciális állományoknak is.

Ahány ismerv-hozzáférésű tételt határoznak meg, annyi invertált index keletkezik. Ezek alapján jelennek meg az online rendszerekben a különféle indexek. (Az indexek valójában mutatók, akár csak a könyvek végén a név- és tárgymutatók. A továbbiakban a felhasználó számára képernyőn megjelenített formát index helyett online mutatónak fogjuk nevezni.)

Az alábbiakban (8. ábra) egy bibliográfiai keresőrendszer választható ismérveinek felsorolása, s vele az online mutatók választéka látható.

A **“keresés”** kifejezést többnyire a pontos kifejezés, mutatónév alapján végzett lekérdezésre tartják fenn, amikor megadunk egy keresőszót, és várjuk, hogy a rendszer megjelenítse a találatokat. Ilyen kereséskor a találatok nemcsak dokumentumtételek, hanem például online mutatók szavai (tehát ismérvek, ismérvtételek) is lehetnek, attól függően, hogy milyen fájlban keresünk.

A **“böngészés”** kifejezést meghatározott ismerv szerint rendezett dokumentumtételek között, meghatározott belépési helyen kezdett keresés. Történhet mutatótételek, tárgyszórendszerek és teauruszok lexikai egységei, osztályozási rendszerek jelzetei között is. Afféle bóklászás, ide-oda keresés, melynek célja az állományon belül az alkalmas tétel, ismerv megtalálása. Néha a **“tallózás”** kifejezést is használják erre a fajta tájékozódásra. Amikor böngészve, tallózva, bóklászva ráakadunk a megfelelő keresőszóra, jelzetre, és megadjuk, hogy e szerint akarjuk lekérdezni a tárolót, akkor ettől a pillanattól kezdve megint **“keresésről”** van szó. A leírt terminológia azonban még nem gyökeresedett meg, a három kifejezést olykor szinonim értelemben is használják.



8. ábra. Online bibliográfiai keresőrendszer mutatóinak választéka.

A példaként bemutatott rendszerben az egyes mutatókon belül **“tallóznak”**, a dokumentumrekordok között **“böngésznek”**, ismert mutatónév alapján pedig **“keresnek”**.

A példában a 637.044 (Kutyák otthoni tartása) ETO-jelzet alapján kezdődik böngészés az ETO-mutatóban. Webhelye: <<http://w3.oszk.hu/hektor.htm>>

Régi információkereső rendszerekben (melyek közül számos még ma is működik) megkülönböztetik a főmutatókat (főindexeket, alapszótárakat) és a mellékmutatókat (kiegészítő szótárakat). Ennek alapja az az elavult megfontolás, hogy vannak a keresés szempontjából fontosabb és kevésbé fontos adatelem-típusok. A főmutatókba kerülnek általában a címek, szerzők, közreadók neve, a tárgyszavak/deszriptorok, az ISBN/ISSN. A mellékmutatókban – ha vannak – kapnak helyet a kiadók, megjelenési helyek, megjelenési évek, a dokumentumok nyelvei stb. Az újabb rendszerekben a két mutatótípust nem különböztetik meg.

Az online keresés logikailag az ismérvek alapján felépülő invertált fájlokra alapszik (a szabad szövegszavakon alapuló keresés másik lehetőség, mellyel a “7.1 Keresés kötött szótár használata nélkül” című fejezetben foglalkozunk).

Az ismérvek lehetnek **szabványosított formájúak és kötetlen formájúak**, származhatnak **kötött szótárból vagy fölvehető szabadon**. Kötött szótár (tárgyszójegyzék, teaurusz, ETO-táblázatok stb.) esetén az ismérv csak abban a formában alkalmazható, ahogy a szótárban szerepel. A szótárak állományát általában meghatározott mezőkhöz (például a tárgyszómezőhöz, nyelvmezőhöz) rendelik. Vannak olyan információkeresésre használt ismérvek is, melyeknek nincs előre megadott, “kötött” szótára, csak szóképzési szabályai. Ezek a személyek, testületek nevei és földrajzi nevek mint besorolási adatok, melyeket az MSZ 3428 szabvány család előírásai szerint kell egységesen leírni (lásd bővebben a 6.3 fejezetet).

Táblázatban:

*

	kötött szótárból	szabadon fölvehető
szabványos formájú	A	B
kötetlen formájú	-	D

1. táblázat. Ismérvek típusai a szabványosítás és a kötöttség szempontjából

*

- A = teauruszok lexikai egységei, tárgyszórendszerek tárgyszavai, osztályozási rendszerek (például ETO) jelzetei, azonosító számok és kódértékek.
- B = személynevek, testületi nevek, címek mint besorolási adatok
- D = “szabad tárgyszavak”

Az invertálás alábbi típusai fordulnak elő:

- **teljes almezőtartalom⁵ invertálása** (phrase indexing) személynevek, testületi nevek, címek mint besorolási adatok, ETO-jelzetei esetében; például “Móricz Zsigmond (1879–1942)” [a HUNMARC 100-as hívójelű mező összes almezőinek együttes értéke];
- szavas invertálás (word indexing) a teljes almezőtartalom egyes szavainak invertálása, címszavak, tartalmi kivonatok szavai esetében; például “Móricz”, “Zsigmond”, “1879”, “1942” [a HUNMARC 100-as hívójelű mező összes almezőinek minden egyes értéke külön-külön]; vagy az “automatizált könyvtári rendszer” tárgyszó esetén az “automatizált”, a “könyvtári” és a “rendszer” (de nem az “automatizált könyvtári rendszer”);
- vegyes invertálás (double posting) elsősorban tárgyszavak/deskriptorok esetén; például keresni lehet az “automatizált könyvtári rendszer” összetett kifejezés, de annak részei, az “automatizált”, a “könyvtári rendszer” és a “rendszer” szerint is.

A címek esetén más mutatóba kerül a teljes almezőtartalom invertálásának eredménye, és más mutatóba a cím szavaiból készült invertálás eredménye (de számos keresőrendszerben a felhasználó kérheti, hogy egyes mutatók összesítve is megjelenjenek).

Az ETO-jelzeteiket általában teljes egészükben invertálják, de a korszerű – egyelőre még csak elvétve előforduló – könyvtári rendszerekben arra is van lehetőség, hogy a kettőspontos összetett jelzeteiket, valamint a különféle általános és korlátozott közös alosztásokat külön is megjelenítsék

⁵ Mivel ma már az adatsere-formátum egyben tárolási formátum is, és mert az adatsere-formátumokban többnyire adott mező almezőjébe kerül adattartalom, ezért beszélünk mindig almezők és nem mezők adattartalmáról (egyszerűbben almezőtartalomról). A(z) (al)mezőknek tartalma, az adatelemeknek értéke van. Az adatelemértékek alkotják a(z) (al)mezők tartalmát.

(azaz vegyes invertálást alkalmazzanak). Az invertálásnak ez a módja az ETO-val végezhető korszerű keresés alapfeltétele.

A tárgyszavakat/deszriptorokat/nemdeszkriptorokat egyes online rendszerekben csak teljes almezőtartalmukkal invertálják, más rendszerekben vegyes invertálást alkalmaznak; egyes online rendszerekben egyetlen mutatóban jelenik meg a teljes almezőtartalom invertálásának és a vegyes invertálásnak az eredménye, másokban kérhető, hogy külön jelenjék meg a teljes invertálás eredményeként keletkező online mutató, és külön a vegyes invertálás online mutatója.

Az ismérvek származhatnak olyan almezőkből, melyek kitöltése kötelező, és olyanokból, melyek kitöltése nem kötelező, csak megengedett (ún. opcionális mezők). Valójában nincs olyan almező, mellyel kapcsolatban ne fordulhatna elő, hogy nincs tartalma. Többkötetes művek egyes köteteinek például hiányozhat a címe, a szerző neve sincs minden mű esetében megadva, hiányozhat a kiadó, a megjelenési hely és év is. A hiányokat és a szerző esetében pótolja, hogy egységesített és szerzőnevet általában kötelező megadni. A megjelenési évet többnyire külön mező pozíciójába is beviszik, hogy kereshető legyen. De például a kiadó, a megjelenési hely vagy az ISBN/ISSN azonosító mezőiben gyakran hiányozhat a tartalom.

Az információkeresés eredményessége szempontjából **nagyon fontos tudni, hogy milyen tulajdonságaik vannak a mezőknek abban az adatbázisban, melyben éppen keresünk**. Ha például a keresésbe bevonjuk a kiadó almező tartalmát is (megadva meghatározott kiadónévet), nagy veszteséggel számolhatunk. A kereséskor nem találjuk meg azokat a dokumentumtétteleket, melyekben hiányzik a kiadó neve (holott a szóban forgó kiadó adta ki a könyvet).

Számítani kell arra is, hogy az adatbázisok szolgáltatói a súgóban (help) nem mindig adják meg pontosan a mezők tulajdonságait, s ezért sokszor csak keserű tapasztalatok alapján ismerhető meg jobban egy-egy adatbázis. Olykor azzal is találkozhatunk, hogy az invertált mezők, horribile dictu az egységesített besorolási adatok állományának karbantartása hiányos, s ezért a belőlük készült online mutatót ellentmondások éktelenítik.

A 9. ábrán a személynevekből készült besorolási adatokból (HUNMARC 100-as hívójelű mező) készült online mutató (böngészőfájl "Móricz V" belépési ponttal kezdődő része) látható.

*



9. ábra. Személynevek besorolási alakjának mutatója. A különböző származású bibliográfiai adatbázisok korábbi egyesítésével keletkezett adatbázisban jól felismerhetők a következetlenségek (például Móricz Zsigmond ötféle, Móricz Virág háromféle besorolási néven is szerepel). Az ilyen mutatókban tehát látványosan megjelennek a feldolgozási hibák. Webhelye: <http://locfind.lib.klte.hu/corvina/opac/wpac.cgi>

3.3 A tiltott szavak

A szavas invertálás alapján keletkező mutatók az adott almezőtartalom minden szavát tartalmazzák, kivéve a leggyakoribb szavakat, amelyekről az a vélemény, hogy nincs információtartalmuk. Ilyen szavak például a kötőszók, névmások, névelők, névutók, de a túl általános jelentésű szavak is (például valami, dolog, jelenség, különbség, időnként). Ezeket a szavakat jegyzékbe foglalják, amelyet negatív szótárnak, irreleváns szótárnak (stoplistának) neveznek. Ezt használják arra, hogy kizárják a benne szereplő szavakat a szavas mutatókban való megjelenítésből. Így a dokumentum (vagy a cím, vagy a tartalmi kivonat) szövegében szereplő összes szónak közel 50 százaléka kizárható anélkül, hogy lényeges információvesztés keletkeznék.

Az Országos Széchényi Könyvtár adatbázisában használt negatív szótár látható a 2. táblázatban. Mivel a nemzeti könyvtárban is különféle nyelvű dokumentumokat dolgoznak föl, a negatív szótár nem csak a magyar nyelvű tiltott szavakat tartalmazza.

a	ko	mellett	v
...	kod	melletti	vaan
alatt	kohtaan	mellől	vagy
alatti	konecno	mert	vai
ale	körül	merthogy	vaikka
alebo	következésképpen	miedzy	vajha
aleer	közé	miatt	vajon
alhoewel	között	mig	valószínű
ali	közötti	mindamellet	van
alig	közül	mint	vanaf
aligha	krome	minthogy	varten
alighanem	kroz	mit	...
all'	...	mitsamt	with
alla	m	mivel	wskutak
...	ma	mivelhogy	wsród
ellen	malgré	múltán	y
ellenben	már	múlva	yinnä
ellenére	másképpen	nyös	z
elől	medu	...	za
előtt	még	u	samiast
en	mégis	über	zatem
ener	meglehetősen	überm	zato
enimvero	mégpedig	übers	zbog
ennélfogva	mégse	...	ze
...	mellé		...

2. táblázat. A tiltott szavaknak az Országos Széchényi Könyvtár adatbázisában használt jegyzéke

3.4 A keresőkérdés megformálása

3.4.1 Logikai műveletek

Amikor keresünk, találatként a keresőszót vagy több keresőszó együttesét tartalmazó dokumentumtöredékek halmazát kapjuk.

Ha a “sportoló” és a “diák” szavakkal kérdezzük, akkor nem elég csak a két szót megadni. Meg kell adni azt a viszonyt is, mely a két keresőszó által reprezentált találati halmaz között fennálljon. A találati halmazok mennyiségi viszonyait (a halmazok egymáshoz viszonyított terjedelmét, illetve helyzetét) logikai műveletekkel állapítjuk meg, e műveleteket pedig logikai műveleti jelekkel – ún. “operátorokkal” – fogalmazhatjuk meg. A megfogalmazás eredményei a logikai kifejezések,

melyek az információkeresésben halmazok közötti műveleteket határoznak meg. Ilyen logikai kifejezésekkel meghatározhatjuk például, hogy megkapjuk mindazokat a dokumentumtétteleket, melyek tartalmazzák

- a “sportoló” és a “diák” szavakat;
- a “sportoló” vagy a “diák” szavakat;
- azokat a dokumentumtétteleket, melyek a “sportoló” szót tartalmazzák, de nem tartalmazzák a “diák” szót;
- azokat a dokumentumokat, melyekben a “sportoló” és a “diák” szó között például legfeljebb 5 szó fordul elő a szövegben;
- azokat a dokumentumtétteleket, melyeknél a megjelenési év értéke például 1960-nál nagyobb.

Többféle logikai műveletet ismerünk (Boole-algebrai műveletek, közelségi műveletek).

Boole-műveletek

- **ÉS** (logikai szorzás, halmazok metszete, konjunkció, AND).

Két vagy több halmaz közös elemeinek meghatározására való.

Például: “marketing” ÉS “bibliográfia”. Ilyenkor találatként megkapjuk mindazokat a dokumentumtétteleket, melyekben ismérvként ez a két szó együttesen előfordul. Ebből azonban nem következik, hogy a két kifejezésnek **egymás mellett** vagy akárcsak **azonos sorrendben** kell előfordulnia dokumentumon belül. Ennélfogva nem biztos, hogy a kapott találatok mindegyike a marketing bibliográfiáját tartalmazza vagy a marketing bibliográfiájáról vagy éppenséggel a bibliográfia marketingjéről szól (az utóbbi problémával kapcsolatban lásd még később az EGYÜTT műveletet); a két kifejezés egymástól értelmileg teljesen függetlenül is előfordulhat egy dokumentumban.

- **VAGY**, megengedő (logikai összeadás, halmazok egyesítése, diszjunkció, OR).

Két vagy több halmaz összes olyan elemének meghatározására való, melyekben az egyik vagy a másik vagy mindkét kifejezés előfordul.

Például: “marketing bibliográfia” VAGY “marketingbibliográfia”. Ilyenkor találatként megkapjuk mindazokat a dokumentumtétteleket, melyekben vagy az egyik vagy a másik kifejezés előfordul, de előfordulhatnak együtt is.

- **VAGY**, kizáró (logikai összeadás, halmazok egyesítése, antivalencia, XOR).

Két vagy több halmaz összes olyan elemének meghatározására való, melyekben vagy csak az egyik vagy csak a másik fordul elő.

Például: “vasúti közlekedés” VAGY “közúti közlekedés”. Ilyenkor találatként megkapjuk mindazokat a dokumentumtétteleket, melyekben vagy az egyik, vagy a másik kifejezés előfordul, de együtt nem fordulhatnak elő. Ezt a műveletet csak azért említjük meg, hogy világos legyen: a VAGY önmagában kétféle műveletet jelent. Ennek például a programozásban jelentősége van, a könyvtári információkeresésben kevésbé. Nagyon különleges eset, ha valakit nem érdekelnek például azok a dokumentumok, melyek mind a vasúti közlekedéssel, mind az közúti közlekedéssel foglalkoznak, hanem csak azok, melyek vagy az egyikkel vagy a másikkal foglalkoznak. Ennek többnyire csak statisztikai szempontból lehet jelentősége.

- **ÉS NEM** (logikai tagadás, halmazok kizárása, non-implikáció, NOT).

Az egyik halmaz elemeiből kizárja a másik halmaz elemeit.

Például: “közlekedés” ÉS NEM “vasúti közlekedés”. Ilyenkor találatként megkapjuk mindazokat a dokumentumtétteleket, melyekben az első kifejezés szerepel, de a második nem.

- **NEM-VAGY** (kizáró logikai tagadás, halmazok kizáró kizárása, Peirce-féle non-implikáció, NOR)

A halmaz(ok) elemeit kizárja az összes többi elem halmazából.

Például: “vasúti közlekedés” NEM–VAGY “közúti közlekedés”. Ilyenkor találatként megkapjuk mindazokat a dokumentumtégeket, melyekben sem az első, sem a második kifejezés nem szerepel. Ezt a műveletet célszerű előválogatott találatok halmazán elvégezni. Például kikeressük az összes közlekedésre vonatkozó tételt a “közlekedés” keresőszóval, és ezek dokumentumtégeiben keressük a “vasúti közlekedés” és “közúti közlekedés” keresőszavakat NEM–VAGY operátorral, miáltal megkaptuk mindazokat a közlekedésről szóló dokumentumokat, melyek nem foglalkoznak se a vasúti, se a közúti közlekedéssel.

Fontos tudni, hogy a fenti eredményre csak akkor jutunk, ha először végrehajtjuk az előválogatást és csak ezt követően használjuk az előválogatott halmazra a NEM–VAGY műveletet. Előválogatás nélkül megadva a “közlekedés” ÉS (“vasúti közlekedés” NEM–VAGY “közúti közlekedés”) keresőképet más eredményhez jutunk.

– VAGY NEM (implikáció, XNOR)

Az egyik halmaz elemeit kizárja az összes többi elem halmazából, kivéve azokat, melyek a másik halmaz elemeivel közösek.

Például: “falkavadászat” VAGY NEM “agár”. Ilyenkor találatként megkapjuk mindazokat a dokumentumtégeket, melyekben az első kifejezés szerepel, és azokat, melyekben a második kifejezés nem szerepel. Ezt a műveletet is célszerű előválogatott találatok halmazán elvégezni. Például kikeressük az összes vadászatra vonatkozó tételt a “vadászat” kifejezéssel, és ezekben keresünk VAGY NEM operátorral, miáltal megkapjuk mindazokat a dokumentumokat, melyek a falkavadászatról szólnak (agarakkal is), és mindazokat, melyek az agarak nélküli vadászatról szólnak. Tehát nem zártunk ki a vadászaton belül minden dokumentumtételt, csak azokat, melyek egyben nem a falkavadászatról is szólnak.

A mindennapi használatban elsősorban az első két művelet és az ÉS NEM fordul elő, mivel többi művelet eredményét ezekkel is meg lehet kapni. Alapértelmezés általában az ÉS–kapcsolat. Csak nagyon kevés rendszer van felkészülve a logikai műveletek teljes választékára.

A VAGY–kapcsolat esetén olykor még a rendszer sűgőja alapján sem egykönnyen állapítható meg, hogy megengedő vagy kizáró VAGY–kapcsolatról van szó. Általában az első a szokásos.

Összehasonlító műveletek (relációs műveletek)

– \< kisebb mint, korábbi mint

Például: “\<2001” esetén minden 2001 előtti dokumentumtétel megjelenik találatként, de az 2001. évben megjelent dokumentumok tételei nem.

– \> nagyobb mint, előbbi mint

Például: “\>2001” esetén minden 2001 utáni dokumentumtétel megjelenik találatként, de az 2001. évben megjelent dokumentumok tételei nem.

– \<= kisebb vagy egyenlő

Például: “\<=2001” esetén minden 2001-ben vagy az előtt megjelent dokumentumtétel megjelenik találatként.

– \>= nagyobb vagy egyenlő

Például: “\>=2001” esetén minden 2001-ben vagy az után megjelent dokumentumtétel megjelenik találatként, beleértve az 2001. évben megjelent dokumentumok tételeit is.

– \<\> nem egyenlő

Például: “\<\>=2001” esetén a 2001. évvel minősített dokumentumtételek nem jelennek meg találatként, az összes többi évvel minősített dokumentumtétel megjelenik találatként.

– = egyenlő

Például: “=2001” esetén csak a 2001. évvel minősített dokumentumtételek jelennek meg találatként, az összes többi évvel minősített dokumentumtétel nem jelenik meg találatként.

Közelségi (távolsági, helyzeti) műveletek

– EGYÜTT két megadott keresőszó között nem fordulhat elő meghatározott számúnál több másik kifejezés, a kifejezések sorrendje számít. Ha például ez az érték 5, akkor a

“fiatalkorú EGYÜTT interjú” esetén találatként megjelenik az a dokumentumtétel, melynek szövegében szerepel a “fiatalkorú kábítószer-fogyasztókkal készült interjú” részlet, de nem jelenik meg az, amelynek szövegében az “interjú azzal, aki fiatalkorú” részlet szerepel. Másik példa: ha az ÉS–kapcsolatnál említett bibliográfia marketingjéről szóló dokumentumokat akarjuk csak megkapni akkor a “bibliográfia EGYÜTT marketing” kifejezést adhatjuk meg, és ha az EGYÜTT értéke megadható a rendszerben, célszerű a 2 megadása, hogy a jelzős szerkezetű marketingkifejezések is relevánsak legyenek.

A távolságot kifejező számérték (példánkban az 5) rendszertől függ. Vannak rendszerek, melyekben a felhasználó adhatja meg az értékét, a rendszerek többségével azonban kötött az érték.

- KÖZEL két megadott keresőszó között nem fordulhat elő meghatározott számúnál több másik kifejezés, a kifejezések sorrendje nem számít. Például a “fiatalkorú KÖZEL interjú” esetén találatként megjelenik az a dokumentumtétel, melynek szövegében szerepel mind a “fiatalkorú kábítószer-fogyasztókkal készült interjú”, mind az “interjú készítő fiatalkorú” részlet.

A távolságot kifejező számértékre az előbb mondottak érvényesek.

3.4.2 Keresés szótöredékekkel és jelhelyettesítéssel

A találati halmaz terjedelmét azzal is befolyásolhatjuk, hogy a keresőszó elejét, végét vagy azon belül meghatározott karakterpozíciókat nyitottá tesszük, azaz megengedjük, hogy a kifejezés elején, végén vagy meghatározott karakterpozícióin bármilyen karakter szerepeljen. Ehhez afféle “jolly joker” jeleket helyezhetünk el a keresőszóban.

- Helyettesítés: Meghatározott karakter (ez rendszerenként változik) bármely karaktert helyettesít. Például a “V#r#svár #” esetén mind a “Veresváry”, mind a “Vörösváry”, mind a “Vöresvári”, mind pedig a “Verösvári” keresőszónak számít.
- Csonkolás: Keresés szótöredékkel. A szó elején (“balról csonkolás”) és/vagy végén (“jobbról csonkolás”) álló meghatározott karakter (ez rendszerenként változik) minden megelőző és/vagy követő karaktert helyettesít. Például a “?hajt?” esetén mind a “meghajtás”, “meghajtó”, “meghajtók”, “meghajtóberendezés”, “meghajtóberendezések”, “meghajtóeszköz”, “meghajtóeszközök”, “meghajtómű”, “meghajtóművek”, “hajtó”, “hajtók”, “hajtóberendezés”, “hajtóberendezések”, “hajtóeszköz”, “hajtóeszközök”, “hajtómű”, “hajtóművek” keresőszónak számít. Keresőszónak fog számítani a “hashajtó” is, amit a NEM művelettel lehet kizárni. A szó elején, “balról” és egyszerre mindkét oldalon csak nagyon kevés rendszerben lehet csonkolni. Számos rendszerben a balról csonkolás se lehetséges.

3.4.3 Keresés összetett kifejezés, illetve meghatározott írásmódú kifejezés szerint

Jelölhető, hogy a megadott több tagú, külön írt szavakból álló kifejezéseket, teljes szövegrészeket (“szóláncokat”) a rendszer egyetlen keresőszóként kezelje. Ha valaki egy magyar nóta teljes szövegét keresi az interneten, akkor például a címét vagy a kezdősorát az AltaVizsla keresőrendszerében idézőjelek közé téve adhatja meg,

“lehullott a rezgő nyárfá aranyszínű levele....”,

és ebben az esetben csak azok a találatok jelennek meg, melyekben az idézőjelek közé tett szövegrész szerepel.

Számos rendszerben nincs írásmód-ellenőrzés, akárhogy adják meg a keresőszavakat, megjelennek mind a kis-, mind a nagybetűvel írt ismérveket tartalmazó találatok. A keresőrendszerek egy

részében jelezhető vagy jelölhető, hogy a keresőszó írásmódját a kereséskor vegyék figyelembe. Jelezhető (például) azzal, hogy a keresőszavakon belül nagybetűt (is) használnak vagy jelölhető (például meghatározott jelre, ikonra való kattintással), hogy a keresőszó írásmódját a kereséskor a rendszer vegye figyelembe. Például csak a “FRÁSZ” vagy a “Frász” írásmódú keresőszó érvényesüljön a kereséskor, nem pedig a “frász” (az első lehet egy rövidítés, a második egy együttes neve).

A közelségi műveletekkel, szótöredékekkel és összetett kifejezéssel végzett kereséssel a 7.1 fejezetben foglalkozunk.

3.4.4 Műveleti jelek

A felsorolt műveletek jelölését nem egyformán szabják meg az egyes online rendszerekben. **Az ÉS–kapcsolatot alapértelmezésben általában a szóközzel (továbbá a +, &, %, \$, /, # és ~ jelekkel) jelölik**, de többnyire megadható az AND vagy az ÉS szavakkal vagy mindkettővel is. A keresőszavak pedig nem szótöredékeknek, tehát például nem jobbról csonkolt kifejezésnek, hanem pontosan megadott kifejezéseknek számítanak.

A többi művelet esetén az angol OR, NOT, NEAR stb., kisebb részükben a magyar VAGY, NEM, EGYÜTT stb. jelek használatosak.

Az online könyvtári rendszerekben többnyire nem kell beírni a műveleti jeleket, hanem a beviteli mező közelében megadják a jelüket, melyekre csak rá kell kattintani. Annak érdekében, hogy egyszerre több különböző kapcsolat is jelölhető legyen, egyszerre több beviteli mezőt is felkínálnak (példaként a “3.6 Keresési stratégia” című fejezet ábráira utalunk).

3.4.5 Keresőkép szerkesztése (profilszerkesztés)

A dokumentum ismérveinek összessége az ismérvlánc. Néhány információkereső rendszerben lehetőség van arra, hogy a dokumentumot több ismérvlánc is reprezentálja.

Az ismérvlánc úgy keletkezik, hogy a dokumentum meghatározott tulajdonságait információkereső nyelvi kifejezésekre fordítjuk. A lehetséges kifejezéseket (az ismérveket) abból a szempontból, hogy szótárból származnak-e vagy sem (használatuk szótárhoz kötött vagy sem) és hogy szabványosítottak-e vagy sem, az 1. táblázatban foglaltuk össze. Ennek alapján például a 6. ábra ismérvei a következő típusokba tartoznak:

Szabványosított, de kötött szótár nélküli:

Móricz Zsigmond (1879–1942);

Iciri piciri; 963–571–418–1;

Bárczi Emese; Handzlik, Jerzy; 1991

Kötött szótári:

087.5(084.1); hu

Kötetlen:

Livroservo

Ezek az ismérvek összességükben a 6. ábra dokumentumtételének ismérvláncát alkotják. Látható, hogy nem minden szabványosított ismérv származik kötött szótárból (de minden kötött szótári ismérv szabványosított). Az első csoporthoz tartozó, szabványosított, de kötött szótár nélküli ismérveket ugyanis meghatározott előírások (például az MSZ 3424 szabványcsalád előírásai) szerint kell képezni, többnyire akkor, amikor első fölhasználásukra sor kerül, az utóbbiakra viszont az jellemző, hogy általában előre szótárba foglalják őket.

Nagyon kevés az olyan rendszer, amelyben az ismérvláncon belüli ismérvek között ún. szintaktikai relációk (szerepjelölők és relátorok) adhatók meg. Ez utóbbiak a dokumentumhoz kapcsolt ismérvek között az adott esetben (szövegekörnyezetben) fennálló összefüggéseket fejezik ki.

Az ismérvlánc (ritkán az ismérvlánccok összessége), az egyes ismérvek minősítése és az ismérvek közötti szintaktikai relációk alkotják a **dokumentumképet** (dokumentumprofil). A többszörös

ismérvláncokat és a szintaktikai relációkat⁶ a továbbiakban nem tárgyaljuk azok ritkasága miatt, ami nem jelenti azt, hogy a jövőben a jelentőségük nem nőhet meg. Az egyre nagyobb mennyiségű információ ugyanis előbb-utóbb kikényszeríti az egyre finomabb dokumentumleírási és keresési módszereket is.

Ahhoz, hogy eredményesen kereshessünk, **a keresőkérdést is ugyanúgy le kell fordítani információkereső nyelvi kifejezésekre (ismérvekre), mint a dokumentumot leíró, a keresés szempontjából fontos tulajdonságokat.** A fordítás eredménye ugyancsak ismérvlánc. A keresőkérdés ismérvlánc, az egyes ismérvek minősítése és a közöttük kifejezett relációk alkotják a keresőképet (keresőprofil).

A keresőkép ismérvei között – szemben a dokumentumképpel – általános a logikai műveleteket kifejező szintaktikai relációk megadása, elsősorban az ÉS meg a VAGY. A “Móricz Zsigmond 1980 után nem Magyarországon megjelent művei” keresőkérdés keresőképe például a következő:

Móricz Zsigmond **ÉS** >1980 **ÉS** **NEM** hu

Ennek a keresőképnek egy gyakorlati kivitelezése látható a 10. ábrán. Az ábrából kiolvasható, hogy a keresőkép fenti formája leegyszerűsítés. Az ábra jobb oldali beviteli mezőiben ugyanis pontosan meg kell adni, milyen értelemben használjuk az egyes ismérveket (“Név”, “Megjelenési év”, “Országnev kód”). Az ismérvek se a dokumentumképen, se a keresőképen belül nem állnak önmagukban, hanem minősítve vannak attól függően, hogy milyen adatelemtípushoz tartoznak. Ilyen típusok például a személynév mint egységesített besorolási adat, az ETO-jelzet, a tárgyszó, a megjelenési év, az ország (ahol a dokumentum megjelent).

A 6. ábrán látható bibliográfiai tétel dokumentumképét például a következőképpen írhatjuk le (az adatelemek szabványos megnevezését dőlt betűkkel írtuk):

Személynév rendszói elem: Móricz; *Személynév egyéb elem:* Zsigmond; *Személynév kronologikus kiegészítő:* (1879–1942); *Főcím:* Iciri piciri; *Kiadó:* Livroservo; ISBN: 963–571–418–1; *További főtétel személynév rendszói eleme:* Bárczi; *További főtétel személynév egyéb eleme:* Emese; *Melléktétel személynév rendszói eleme:* Handzlik; *Melléktétel személynév egyéb eleme:* Jerzy; *ETO-jelzet:* 087.5(084.1); *Megjelenési év kereshetően megadott bibliográfiai adat:* 1991; *Ország:* hu

Az adatelemek típusát (fajtáját) a könyvtári rendszerek katalogizálási moduljában (ma már az esetek nagy többségében) az adatcsere-formátumok tartalomjelölői minősítik (a mező hívójele, az almező-azonosító stb.). Magyarországon a magyar HUNMARC és az amerikai USMARC adatcsere-formátumok terjedtek el; a két formátum majdnem azonos. A HUNMARC adatcsere-formátum tartalomjelölőit használva a dokumentumkép a következőképpen fest:

100a: Móricz; *100j:* Zsigmond; *100d:* (1879–1942); *240a:* Iciri piciri; *260b:* Livroservo; *020:* 963–571–418–1; *700a:* Bárczi; *700j:* Emese; *700a:* Handzlik; *700j:* Jerzy; *080a:* 087.5(084.1); *008 7–10:* 1991; *008 15–16:* hu

Az online információkereső rendszerekben (a könyvtári nyilvános online katalógusokban is) az ismérvek minősítésére különböző módszereket alkalmaznak (lásd a 6. ábra példáját). A példaként megadott ábrában a HUNMARC tartalomjelölőit használták. Más esetekben két-három karakteres kódokkal minősítik a keresőkép ismérveit (olykor prefix, illetve suffix a nevük, attól függően, hogy a rendszer szabályai szerint az ismérv előtt vagy után kell állniuk).

A keresőnek nem kell ismernie a tartalomjelölőket, de a keresőprogramnak igen (a – laikus – keresőnek arról sem kell tudnia, hogy mi fán terem a dokumentum- meg a keresőkép).

A keresőnek csak azt kell tudnia, a laikus keresőnek pedig arra kell rájönnie, milyen típusú adatelem az általa választott keresőszó. Amikor mondjuk Móricz Zsigmond 1991-ben Magyarországon megjelent műveit keresi, tehát személynév, ország és évszám szerinti keresést

⁶ Szöveggyűjtemény, o. 194, Ungváry Rudolf: Az osztályozás alapjai, p. 74

végez egyidejűleg, akkor például az Országos Széchényi Könyvtár keresőfelületén fel kell ismernie, hogy a “Név”, a “Megjelenési év” és az “Országnev kód” minősítéseket kell megadnia (lásd 10. ábrát). A program a tárolt dokumentumképekben az ennek megfelelő 100-as hívójel a és d almezőjével és a 008 hívójelű mező 7–10., illetve 15-16. pozíciójával azonosított személynév, évszám és országnévkód adatelemek (ismérvek) között végez keresést.

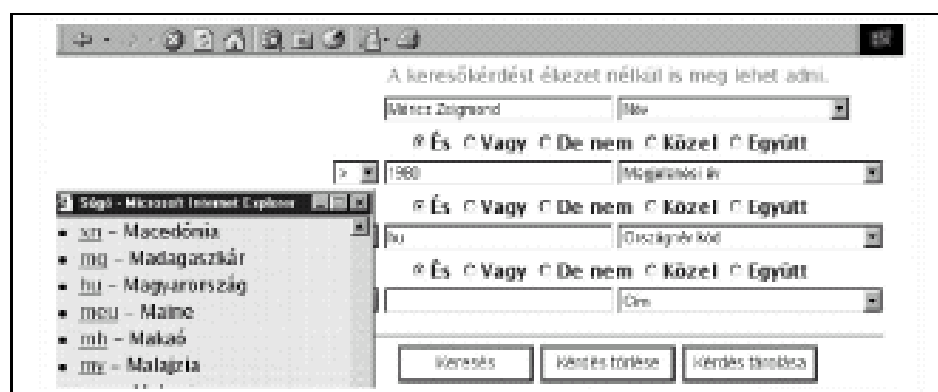
A 10. ábrán látható keresőkérdés pontos keresőképe tehát elvileg a következő:

100a: Móricz 100j: Zsigmond ÉS > 008 7–10: 1980 ÉS NEM 008 15–16: hu

Az OSZK rendszerében a kérdés a “Kérdés tárolása” paranccsal elmentve a következőképpen jelenítik meg:

Móricz Zsigmond (NE) + 1980 (DAT) – hu (FNE)

*



10. ábra. Személynév, megjelenési év és ország szerinti keresés az OSZK keresőfelületén. A jobb oldali beviteli mezőkben választható ki a keresőszó (ismérv, adatelem) típusa (a típusok választékát a jobbszélső legördülő sáv itt nem megjelenített jegyzékéből választhatjuk ki. A bal oldali beviteli mezőbe írhatók a keresőszavak, bal szélükön az összehasonlító műveleti jelek legördülő sávja látható.

A kép bal alsó sarkában az országnévkódok legördülő jegyzéke látható, melynek alapján a kereső kiválaszthatja az országnévkódok kötött szótárából a megfelelő kódszót (példánkban ez a “hu”). Webhelye: <http://www.oszk.hu/>

*

Az egyes ismérvtípusok szerinti keresés speciális kérdéseivel a későbbi fejezetekben foglalkozunk.

3.5 A parancsnyelv

3.5.1 A parancsnyelv szavai

A keresési stratégia és taktika végrehajtása során az információkereső nyelv szavaival és a műveleti jelekkel megfogalmazott keresőképet, továbbá a stratégia és a taktika egyéb lépéseit parancsokkal “adjuk tudtára” a információkereső rendszernek. Az adott információkereső rendszeren belüli lehetséges parancsok (a parancsszavak) összességét és azok használatának szabályait nevezik parancsnyelvnek. Az egyes rendszerekben használt parancsnyelvek kisebb-nagyobb mértékben különböznek egymástól.

A parancsok vonatkoznak az adatbázisba való belépésre, a keresés végrehajtására, a mutató, illetve a rekordok megjelenítésére, a duplumok (ugyanazon tételek többszörös találati előfordulásának) kezelésére, a relevanciavizsgálatra, a kérdések (a keresési történet) mentésére stb.

Parancsként értelmezi az információkereső rendszer a keresőképben meg-adott műveleti jeleket is.

A 3. táblázatban az egyik legismertebb online adatbázis-szolgáltató, a DIALOG fontosabb parancsszavait tüntettük fel példaként.

*

Tartalma	Neve	Jele a keresőképben belül
belépés adatbázisba	BEGIN	B
keresés	SELECT	S
keresés Boole-műveletek nélkül	FIND	F
egyszerűsített keresés (kapcsolatok nélkül); relevancia sorrendben megjelenített találatok	TARGET	T
mutató megjelenítése meghatározott kifejezésnél	EXPAND	E
ÉS	AND	AND
VAGY megengedő	OR	OR
ÉS NEM	NOT	NOT
csonkolás tetszőleges számú karakterre	truncation open	?
csonkolás egy karakterre	restricted	?szóköz?
csonkolás adott számú karakterre (pl. max. 3 karakter hosszban)		???
karakter helyettesítés szóon belül: kérdőjelként egy karakter (pl. 2 karakter hosszban)	internal	??
együtt (adott sorrendben egymás után)	WITH	(W)
együtt (adott sorrendben n szónyi távolságban, pl. 2W)	nW	(nW)
közel egymás mellett tetszőleges sorrendben	NEAR	(N)
közel egymás mellett tetszőleges sorrendben n szónyi távolságban, pl. 5N)	nN	(nN)
a keresőszó legyen egy almezőn/mondaton/paragrafuson belül	SUBFIELD	SUBFIELD
a keresőszó legyen egy mezőn belül	FIELD	FIELD
megjelenítés relevancia sorrendben	RANK	RANK
keresőelemek kiemelése, melyekkel a keresést majd folytatjuk	MAP	MAP
duplumok kiszűrése (több adatbázisban végzett kereséskor)	REMOVE DUPLICATE	RD
összes duplum megkeresése (több adatbázisban végzett kereséskor), csoportosítás egy halmazban	IDENTIFY DUPLICATES	ID
ismétlődésenként csak egy-egy duplum megjelenítése (több adatbázisban végzett kereséskor), egy csoportosítása egy halmazban	IDENTIFY DUPLICATES ONLY	IDO
keresőkép ideiglenes mentése	SAVE TEMP	SAVE TEMP
ideiglenesen elmentett keresőkép előhívása	RECALL	RECALL
előhívott keresőkép újrafuttatása	EXECUTE	EX
keresési történet felidézése	DISPLAY SETS	DS
találatmegjelenítés, nyomtatás	TYPE, PRINT	T, PR
kijelentkezés	LOGOFF	LOGOFF

3.táblázat A DIALOG parancsnyelve, webhelye:

<http://library.dialog.com/success/searching.html#command>

*

Amikor a DIALOG rendszerben például a keresésre a “select” parancsot adják, akkor a program elvégzi a parancs után következő keresőkép összehasonlítását a tárolt dokumentumtételek ismérveivel. Például

select toxicitás (W) vizsgálat??

azt jelenti, hogy a “toxicitás” és a “vizsgálat??” közvetlenül egymás után forduljanak elő, és a “vizsgálat” keresőszó után legfeljebb két karakter állhat (például “vizsgálatok”). A (W) elé együttható (numerikus prefix) tehető, mellyel az ismérvek között nagyobb távolságot adhatunk meg.
A

select toxicitás (2W) vizsgálat??

azt jelenti, hogy a “toxicitás” és a “vizsgálat??” között legfeljebb két szó fordulhat elő a beírás sorrendjében, azaz relevánsak azok a dokumentumtételek, melyekben például a “toxicitásvizsgálat” és a “toxicitás- és környezetvédelmi vizsgálat” kifejezések fordulnak elő.

A parancsok választékánál nehezebben tekinthetők át a parancsok használatának szabályai. Az eredményes kereséshez nagyon sok részletet kell ismerni. A jelek használatához afféle “fogások” is fűződnek. Két példával érzékeltetjük a felvetődő problémákat (ezek azonban csak a jéghegy csúcsai).

1. példa

Az “és” szó a legtöbb keresőrendszerben tiltott kifejezés. Ha a kérdés a “szülők és gyerekek” kapcsolatával foglalkozó irodalom, akkor a pontosabb találatokhoz érdemes nem csak külön a “szülő” és a “gyermek”, hanem a “szülők és gyermekek” kérdést is feltenni. Ha az utóbbit csonkolva adjuk meg keresőkérdésnek (e példában magyar műveleti jeleket használva)

szülő? és gyer?ek?

akkor a rendszer figyelmen kívül fogja hagyni az “és” kötőszót, és mert a szóköz a legtöbb rendszerben alapértelmezésben az ÉS–kapcsolat jele, csak azokat a találatokat jeleníti meg, melyeket a

szülő? ÉS gyer?ek?

keresőképpel adtak meg. Az összetett kifejezéssel végzett keresés érdekében idézőjelek közé téve kell megadni a keresőszavakat a megfelelő logikai kapcsolatokkal. Az idézőjelek közé tett kifejezést általában a legtöbb rendszerben egyetlen keresőszónak értelmezi a program:

“szülők és gyermekek” **VAGY** “szülő és gyermek” **VAGY** “szülő és gyermekek” **VAGY** “szülők és gyerekek” **VAGY** “szülő és gyerekek” **VAGY** “szülő és gyerek”

2. példa

Ha például a Frankensteinről és Draculáról szóló témában keresünk az interneten, és az AltaVizsla keresőrendszerében megadjuk az alábbi ismérvláncot

frankenstein dracula

akkor ezt az AltaVizsla keresőrendszere ÉS–kapcsolatként értelmezi. Ugyanazokat a találatokat eredményezi, ha a szóban forgó rendszerben a keresőszavak előtt + jellel jelezzük ezt a kapcsolatot:

+ frankenstein + dracula

Mindkét esetben megkapjuk mind a kis, mind a nagy (kezdő)betűkkel írt ismérveket tartalmazó találatokat.

Ha viszont akárcsak egyetlen keresőszóban nagy (kezdő)betűt használunk, például

frankenstein Dracula

akkor már csak azok a találatok jelennek meg, melyekben az első keresőszó kis kezdőbetűvel, a második pedig nagy kezdőbetűvel szerepel ismérvként. Ha tehát a keresőrendszer nagy betűt is talál a keresőszavakban, akkor automatikusan áttér az írásmód ellenőrzésre.

Ha az ÉS kapcsolat műveleti jelét csak az egyik keresőszó elé tesszük ki

frankenstein + dracula

akkor ez azokat a találatokat eredményezi, melyekben a “dracula” (és a “Dracula” meg a “DRACULA” stb.) feltétlenül előfordul, a “frankenstein” (és a “Frankenstein” meg a “FRANKENSTEIN” stb.) pedig vagy előfordul, vagy nem.

3.5.2 Parancssal vezérelt és felülettel vezérelt keresés

A közvetlen parancssal vezérelt kereséskor (command search, command driven search) a keresőnek ismernie kell a parancsokat, a keresőképnek és a keresési stratégiának az adott rendszerben alkalmazott szabályait, és lényegében minden közvetlen támogatás nélkül, a képernyő jóformán üres felületén kell beírnia a rendszer parancsnyelvén a kérdéseket és megadni a találatok rendezésére és megjelenítésére stb. vonatkozó utasításait (ilyen parancsnyelv példáját láthattuk a 3. táblázatban). Történelmileg ez volt a korábbi eljárás és máig megőrizte jelentőségét. A parancssal vezérelt kereséshez gyakorlat kell, részleteiben jól kell ismerni az adott keresőszolgáltatás keresési lehetőségeit még alapfokon is.

Ez a keresésvezérlés **a távoli hozzáférésű kereskedelmi adatbázis-szolgáltatásokat jellemzi.** A rendkívül gazdag parancsnyelvi választékról a szolgáltatások sűgője tájékoztat, és többnyire betanító programok (tutorials) is rendelkezésre állnak. A sűgőkban egyrészt megtalálható mindazoknak a műveleteknek, parancsoknak, eljárásoknak a magyarázata, melyek a rendszerben használhatók, másrészt használatuk példái, keresési példák. A nagy rendszerek sűgői is nagyok, több fejezetet és alfejezetet tartalmaznak. A gyakorlatlan keresőnek a sűgőkban is nehéz eligazodnia. A DIALOG rendszerben például a mezőazonosító kódok választéka, melyekkel a keresőképben megadható, hogy melyik ismérvet melyik mezőben keresse a rendszer, nem a parancsok sűgőjában, hanem az adatbázis ismertetőjében található. Ha abból indulunk ki, hogy a mezőazonosító kódok az adatbázis szerkezetét jellemzik, akkor ez logikus, a kódok választéka az adatbázis ismertetőjébe való; ha viszont abból, hogy a keresőkép összeállítása szempontjából ezek is afféle keresési parancsok, akkor e kódok a parancsok sűgőjába valók lennének.

A parancssal vezérelt keresés használatának egyik oka, hogy **ezzel lehet a legjobban, a legrugalmasabban és a leggyorsabban kiaknázni a keresőrendszer adta lehetőségeket** – feltéve, ha már ismerjük a parancsnyelv minden csínját-bínját. A legösszetettebb kérdések, a legbonyolultabb elérési utak és a legrészletesebben meghatározott rendezési és megjelenítési igények teljesen kötetlenül, egyetlen összetett keresőprofilban adhatók meg anélkül, hogy a rendszeren belül tájékozódás vagy adatbevitel céljából különböző helyekre kellene ugrálni.

Használatának másik, történelmi oka anyagi természetű. A kereskedelmi szolgáltatásokban végzett keresés minden másodperce pénzbe kerül. Parancssal vezérelt keresés esetén – a parancsok és használatuk előzetes, alapos ismeretében – föl lehet készülni a keresési stratégiára, és úgy alakítható ki a teljes keresőkép, hogy be se léptünk még a rendszerbe. A belépésre csak ezután kerül sor, és azonnal bevihető a kérdés, mely már azokat az utasításokat is tartalmazza, hogy milyen formában adja ki (nyomtassa, faxolja stb.) a rendszer a találatokat. Ezáltal **lényegesen csökkenthető a rendszer használatának az ideje, s vele a keresési költségek.** (Ma ennek már kisebb a szerepe, mert csökkentek az idő- és növekedtek a használati díjak.)

A felülettel vezérelt kereséskor (guided search, menu driven search) a keresőnek nem kell előre ismernie a konkrét parancsszavakat, a keresőképnek és a keresési stratégiának az adott rendszerben alkalmazott részletes szabályait, nem kell különösebb gyakorlat – legalábbis alapfokon. Az egyes adatelemtípusokhoz vagy előre megadják a beviteli mezőket, vagy a mezőkhöz rendelt legördíthető mutatókból választható ki az adatelem típusa, vagy mindkettőt vegyesen alkalmazzák. A műveleti jelek vagy egy részük is legördíthető mutatókból választható ki. Az átfogóbb parancsok számára (keresés, kiválasztás, rendezés, nyomtatás stb.) külön parancsgombok állnak rendelkezésre a kezelőfelületen. **Ezzel a vezérléssel lehet a legtöbb közvetlen segítséget adni a felhasználónak.** Történelmileg ez volt a később kialakult eljárás – a CD-ROM adatbázisok megjelenésével párhuzamosan alakult ki –, és ma egyre nagyobb teret hódít.

Sokoldalú rendszerek esetében minden keresési lehetőséget képtelenség egyetlen felületbe beleprésselni. A különböző beviteli mezőkhöz, legördíthető mutatókhoz, parancsgombokhoz stb.

hely kell. Ráadásul fontos szempont, hogy minden a lehető legegyszerűbb legyen, nehogy elriasszák a kezdő felhasználókat. Ezért az ilyen vezérlések tervezésekor dönteni kell arról, mi legyen éppen a legfontosabb (pontosabban: miről feltételezik, hogy az egyes felhasználói csoportok számára a legfontosabb), és a vezérlést ennek alapján alakítják ki.

Mivel szinte minden lehet valamilyen nézőpontból vagy valamelyik felhasználói csoport számára fontos, ezért nagyon sok különböző vezérlőfelületet kell tervezni. Külön-külön ezek mindegyike nagyon egyszerű, együtt mégis nagyon bonyolult szerkezetet alkothatnak.

Az integrált könyvtári rendszerek nyilvános online katalógusában a keresés általában ingyenes. Itt nem elsősorban a keresés költségei okoznak problémát, mint inkább az, hogy számítani kell a teljesen gyakorlatlan keresőkre. Ezért ezekben a rendszerekben csak néhány meghatározott esetben fordul elő parancsmódú keresés. Jellemző, hogy igyekeznek a tájékozatlanabb felhasználó igényeinek is megfelelő felülettel vezérelt keresési lehetőségeket biztosítani. Legfeljebb a logikai, közelségi stb. műveleti jeleket kell parancsként bevinni a keresőszavak közé, de a szóköz alapértelmezésben a leggyakrabban használt ÉS–kapcsolatot jelenti, és ezért a logikai műveleti jel hiányában is használható a rendszer.

Az internet esetében a mondottak hatványozottan érvényesek. Az internet World Wide Web (WWW, “világháló”, a továbbiakban web) felülete eleve abból a célból keletkezett, hogy még azok is használják az internetet, akik egyébként könyvtári rendszereket sem vesznek igénybe. Az internet körülményei között a HTML-dokumentumokon (weblapokon) túlmenő keresés, így a könyvtári dokumentumtétel (a bibliográfiai információ) keresése pusztán abból a szempontból, hogy mire használják az internetet, ma (bár összességében jelentős, de a web egészéhez képest még) alárendelt jelentőségű. Portálszolgáltatások, honlapok, hírek, csevegővonalak stb. keresése az uralkodó. Noha az internetes keresőszolgáltatások olykor nagyon színvonalas részletes keresési lehetőségeket (powered search, advanced search) is föl kínálnak, ezekben is túlnyomórészt felülettel vezérelt a keresés.

A könyvtári és internetes felülettel vezérelt keresőrendszerekben a parancsok és a vezérlőfelületek választéka többnyire (és talán csak egyelőre) nem is olyan gazdag, mint a paranccsal vezérelt rendszerekben. Több felületről azonban kattintással már elérhető a szakképzett keresők rendelkezésére álló, nemzetközileg szabványosított CCL (Common Command Language) általános parancsnyelv alkalmazását lehetővé tévő keresőfelület.

Mindebből következik, hogy a felülettel vezérlés esetében a keresés nemcsak lassabb, hanem a paranccsal vezérelt keresés szabadságfokához képest többnyire szegényesebb. Probléma azért nem kerekedik belőle, mert egyrészt a könyvtári rendszerek és az Internet használata (ha az adatátviteli vonalak – a telefonálás – használati költségeitől eltekintünk) **ingyenes**. A nem fizető felhasználó “ráér”. Másrészt a **használatuk – első megközelítésben – könnyű**. A felhasználónak nem kell különösebb szellemi erőfeszítéseket tennie. Harmadrészt a keresők túlnyomó részének az igényeit a felülettel vezérelt keresés kielégíti.

Mindebből azonban nem következik, hogy ha a felhasználó maradéktalanul ki akarja meríteni egy könyvtári felülettel vezérelt rendszer (de akár egy internetes keresőszolgáltatás) lehetőségeit, akkor ez nagyon egyszerű volna. Az igényes kereséshez ezekben a rendszerekben is hozzáértésre van szükség, annál is inkább, mert ezekben a rendszerekben is egyre gyakoribb a fejlettebb információkereső nyelvek (tezauruszok, átfogó hierarchikus osztályozási rendszerek) használata.

Noha **ez a vezérlési mód a könyvtári és internetes rendszereket jellemzi**, a kereskedelmi szolgáltatásokban (mind a távoli hozzáférésű, mind a CD-ROM-szolgáltatásokban) is teljes gazdagságukban megtalálhatók. Szemben a könyvtári és internetes rendszerekkel ráadásul sokkal nagyobb gazdagságban, s ennél fogva bonyolultságban is. Ennek oka, hogy a kereskedelmi szolgáltatók is meg akarják nyerni maguknak a laikus keresőket (történelmileg ugyanis a professzionális keresők világa volt ez a terület, ahová a laikusok csak az internet megjelenésével juthattak el).

A DIALOG rendszerben végzett paranccsal vezérelt keresés keretében például az előzőkben említett és a 10. ábrán az OSZK keresőrendszerében megjelenített

100a: Móricz 100j: Zsigmond ÉS > 008 7–10: 1980 ÉS NEM 008 15–16: hu

keresőkép a 11. ábra alsó részében látható.

*

```
DialogClassic Web (tm)
B.426
03jan02 06:18:50 User209681 Session D3.2
$1.68 0.525 DialUnits File426
$0.00 4 Type(s) in Format 6
$0.00 4 Types
$1.68 Estimated cost File426
$0.79 INTERNET
$2.47 Estimated cost this search
$2.54 Estimated total session cost 0.721 DialUnits

File 426:LCMARC-Books 1968-2001/Dec W4
(c) format only 2001 Dialog Corporation

Set Items Description
---
?
S AU=MORICZ, ZSIGMOND AND PY>1980 NOT CP=HUNGARY
 32 AU=MORICZ, ZSIGMOND
3822202 PY>1980
40519 GL=HUNGARY
S1 0 AU=MORICZ, ZSIGMOND AND PY>1980 NOT CP=HUNGARY
```

11. ábra. Parancssal vezérelt keresés a DIALOG rendszerben. AU a szerző, PY a megjelenési év, CP a megjelenési hely mezők azonosítója. Az elején az addig végrehajtott keresés (session, “keresési ülés”) idejéről és költségeiről látható tájékoztatás. Webhelye: <http://www.dialogclassic.com/>

*

A 11. ábra felső részében az első adat az utolsó keresési művelet (“ülés”, session) dátuma, a kereső azonosítója, az addigi keresés költségei, ez alatt az adatbázis adatai (File 426:LCMARC-Books 1968-2001/Dec W4) láthatók, melyben a keresést végezték (a Kongresszusi Könyvtár könyv-adatbázisában végezték a keresést). A második sorban látható a keresőkép, alatta pedig a találatok száma ismérvenként és az utolsó sorban a teljes keresőképre vonatkozóan, amely nulla, azaz nincs adatbázisban Móricztól 1980 után külföldön megjelent könyv.

A keresés sikere azonban nem csak a kérdésnek megfelelően fogalmazott keresőképen múlik. **Ismerni kell az adatbázis tulajdonságait is, amelyben keresünk.** A Kongresszusi Könyvtár GL (földrajzi hely) mezőjét például nem kötelező kitölteni (ún. opcionális mező). Ezt a mezőt a mű tartalmára vonatkoztatva használják: azt a földrajzi helyet jelenti, amelyről a mű szól. A keresőprogram a NEM parancs esetén az üres mezőt is úgy értelmezi, hogy a mű nem az adott országról szól. Holott lehetséges, hogy róla szól, de mert nem kötelező kitölteni, ezért előfordulhat, hogy akkor se töltik ki, amikor indokolt lett volna. Ezért a NEM GL=hungary paranccsal végzett keresés eredményeként olyan találatokat kapunk, melyek dokumentumai Magyarországról szólnak ugyan, csak nincs kitöltve a dokumentumtételben a GL mező.

A keresési folyamat során megadott keresőképeket a rendszerek megőrzik. Ha már van legalább egy lekérdezett keresőkép (legalább egy sor), a következő kereséskor az előző sor beépíthető a második keresőképbe (a második sorba).

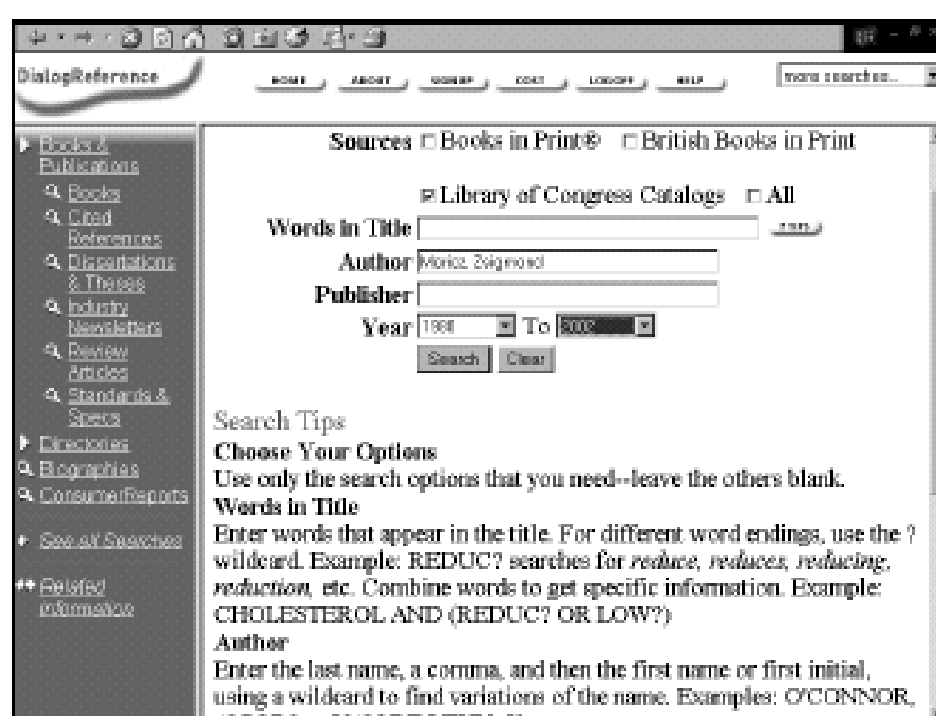
Szemléltetésül az alábbi keresési példát nem egyetlen, egyszerre megfogalmazott keresőképpel mutatjuk be (a folyamatot a DIALOG jelöléseivel, de leegyszerűsítve mutatjuk be):

```
S1 AU=MORICZ, ZSIGMOND AND PY>1980
S2 S1 NOT LA=HUNGARIAN
S3 S2 CP=HUNGARY
```

A könyvtári rendszerekben az egyes keresőképeket (sorokat) általában “keresési történet” vagy ehhez hasonló néven lehet elmenteni, de a sorok összekapcsolása (a fenti példában az S2 S1 NOT LA=HUNGARIAN, illetve az S3+ S2 CP=HUNGARY) a többségükben nem lehetséges. A fenti példát ezekben a rendszerekben a felülettel vezérelt keresés lehetőségeit igénybe véve kell megoldani.

A 12. ábrán ugyanennek a témának egyik felülettel vezérelt keresési módja látható a DIALOGSELECT DialogReference részrendszerben, amely elsősorban bibliográfiai stb., dokumentumtípusok keresésére való. A felületet a könnyű használhatóság érdekében a lehető legegyszerűbbre tervezték, az ilyen típusú keresésekhez legszükségesebb adatelemtípusok figyelembevételével. Noha a példaként választott keresőkérdés nem volt valami bonyolult, de az országnév bevonásával egy kicsit már eltért az átlagos kérdésektől, így egy átlagos kérdésekre tervezett felületen nem kérdezhető le. Ha valaki mégse akar lemondani róla, kénytelen a paranccsal vezérelt keresést választani.

*

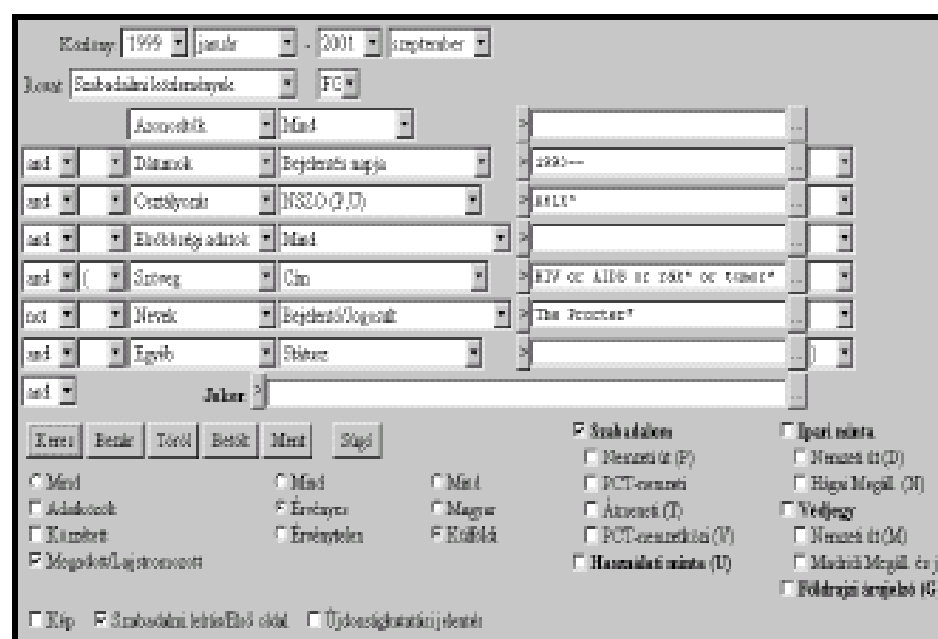


12. ábra. Felülettel vezérelt keresés a DIALOG rendszerben. A “wildcard” tetszőleges karakter használatát jelenti (például ? alkalmazásával). Webhelye: <http://www.dialogselect.com/ref/>

*

A speciális tárgykörű adatbázisok felülettel vezérelt keresőfelületeiben néha igyekeznek minden lehetőségre fölkészülni, amivel e vezérlési mód lehetőségeinek a határáig jutnak el. Erre példa a 13. ábrán látható felület.

*



13. ábra. Iparjogvédelmi adatbázis (PIPACS, Publikus IParjogvédelmi AdatbázisCSalád) keresőfelülete. Keressük mindazokat az érvényes, lajstromozott és teljes leírással is rendelkező szabadalmi dokumentumokat, amelyeket 1993. után a The Procter & Gamble cég kivételével külföldi bejelentők tettek az AIDS, illetve a rák elleni gyógyszerkészítmények területén. A * a jobb oldali csonkolás jele, az A61k a Nemzetközi Szabadalmi Osztályozási rendszernek a keresés tárgyára vonatkozó jelzete, a példában csonkolási jellel megadva, hogy az alárendeltjeit is megkapjuk. (A találatok száma 2002. 01. 03-án 2 volt.)
 Webhelye: <<http://pipacs.hpo.hu/>>, ill. <http://pipacs.arcanum.hu/>
 *

A 14. ábrán összetett paranccsal vezérelt keresést mutatunk be vázlatosan a DIALOG rendszerben (a bevitel félkövéren szerepel és csak a legfontosabb rendszerüzenetek láthatók).

Kérdés: Kik a vezetők, kutatók és kutatólaboratóriumok a celluláris idegrendszeri (celluláris neurális) hálózatok (cellular neural networks) témában?

b411	belépés a 411-es számú DIALINDEX adatbázisba, mely az összes adatbázis mutatóit tartalmazza
sf allsci	keresés az összes természet- és alkalmazott tudományi adatbázisban (fájlban)
You have 135 files in your file list	135 fájl szerepel a jegyzékben
ss cellular(w)neural(w)network?	keressük azokat a dokumentumtégeket, melyekben a három ismerv egymás után fordul elő, a "network" után tetszés szerinti karakterek állhatnak (a találati listán itt nem mutatjuk)
15 files have one or more items	15 fájlban található egy vagy több találat
rank files	a találatok a fájlszámok sorrendjében jelenjenek meg a fájlok a találatok száma szerinti sorrendben jelennek meg
save temp	ideiglenesen mentjük el a keresőképet
Temp SearchSave "TD007" stored	a keresőkép a TD007 azonosító alatt lett elmentve
b N1-N4	belépünk abba a 4 fájlba, amelyben a legtöbb a találat
exs	hajtsuk végre a keresést a mentett keresőképpel
S1 156449 CELLULAR	a találatok jegyzéke ismervenként és találatok a teljes keresőképpel
S2 111008 NEURAL	
S3 430719 NETWORK?	
S4 311 CELLULAR(W)NEURAL(W)NETWORK?	
rank s4 au	az s4 halmaz találatait rendezzük a legtöbb közleménnyel rendelkező szerzők szerint
1 45 CHUA LO	a szerzők relevancia szerinti jegyzékében a 2. oszlop a relevancia mértékét jelenti (a 288 tételből álló jegyzék első 8 tétele látható)
2 35 ROSKA T	
3 32 CHUA, L.O.	
4 27 ROSKA, T.	
5 20 NOSSEK JA	
6 17 CHUA, LEON O.	
7 16 NOSSEK, J.A.	
8 9 NOSSEK, JOSEF A.	
stb.	
1,2,3,4,6,10	mentjük el az első négy, a hatodik és a tizedik találati halmazt (a közbenső lépéseket nem
Saved term(s): 1-4,6,10	találati halmazt (a közbenső lépéseket nem

10;exit;Y
 Creating temporary SearchSave "TD008"
 exs
 elmentett

S5 300 AU="CHUA LO"
 S6 57 AU="ROSKA T"
 S7 308 AU="CHUA L.O."
 S8 54 AU="ROSKA, T."
 S9 168 AU="CHUA, LEON O."
 S10 88 AU="CHUA L.O."
 S11 929 S5:S10

ds

S1 156449 CELLULAR
 S2 111008 NEURAL
 S3 430719 NETWORK?
 S4 311 CELLULAR(W)NEURAL(W)NETWORK?
 S5 300 AU="CHUA L.O."
 S6 57 AU="ROSKA T"
 S7 308 AU="CHUA L.O."
 S8 54 AU="ROSKA, T."
 S9 168 AU="CHUA, LEON O."
 S10 88 AU="CHUA L.O."
 S11 929 S5:S10

ss s4 and s11

S12 143 S4 AND S11

rd s12

S13 55 RD S12 (unique items)

t/3,k/1-2 from each

show files

logoff hold

mutatjuk be)
 kilépés a RANK parancsból
 a tételek a TD007 azonosító alatt lettek elmentve
 hajtsuk végre a TD007 azonosító alatt

ismérvek szerint a keresést lépésenként
 a keresés találati jegyzéke

a szerzők összes művére keresve több találatot
 kapunk, mint a RANK parancs után, ott ui. csak
 az S4 "cellular(w)neural(w)network?" halmazon
 belül kerestünk

jelenítsük meg a keresési történetet

az s4 és s11 halmazok közös részét keressük,
 vagyis a két szerzőnek az adott témában írt
 műveire szűkítjük a keresést

szűrjük ki a duplumokat az S12 eredményből
 (mivel több adatbázisban kerestünk, számítani
 lehet ugyanannak a rekordnak többszörös
 előfordulására)

55 találat maradt
 minden fájlból az első két találatot jelenítjük
 meg (a találatokat itt nem mutatjuk be)
 jelenítsük meg a fájlokat, melyekből a találatok
 származtak (itt nem mutatjuk be)
 kijelentkezés

14. ábra Összetett paranccsal vezérelt keresés a DIALOG rendszerben. A folyamatban a RANK paranccsal relevancia-visszacsatoláson alapuló stratégiát alkalmaztak (lásd erről a 3.6.2.7 fejezetet).

3.6 A keresési stratégia

3.6.1 Automatizált és emberi közreműködéssel végzett keresés

Az információkeresés egyes folyamatainak automatizált eljárásai az információkeresés teljesen formalizált végletét képviselik.

Egyik fajtájuk az automatizált indexelés. Ezen olyan eljárásokat értünk, melyekben a szöveg, referátum vagy a cím szavait gépi segítséggel kivonatolják (extrahálják) akár úgy, ahogy a szavak az eredeti dokumentumban szerepelnek, akár szótövek alapján összevetve (ezt konflálásnak nevezik), akár súlyozva, statisztikai vagy valószínűségi alapon, akár pedig a felsorolt főbb eljárásokat kombinálva.

Másik fajtájuk az automatikus dokumentumosztályozás. Ebben a szöveg szavait arra használják föl, hogy megfelelő matematikai eljárásokkal (például klaszterelemzéssel) hasonlóságokat állapítsanak meg a dokumentumok között. A hasonlóság megállapításának az alapja pedig az, hogy milyen mértékben fordulnak elő a dokumentumokban vagy azok reprezentációiban (például a referátumokban, deskriptorláncokban) közös szavak. E mérték, mint afféle számszerűsített hasonlósági “asszociáció” alapján a dokumentumok csoportosíthatók, és e csoportosítás (klaszterálás) az osztályozás eredménye.

Mindkét eljárás eredményeként általában gépi indexek, online mutatók készülnek, melyek alapján a kereső tájékozódhat.

A természetes nyelvű szövegeket azonban a szintaktikai szerkezet és a szemantikai tulajdonságok jellemzik. Ezekkel a szerkezeti és tartalmi (jelentésbeli) jellemzőkkel az automatikus indexelés és osztályozás művelői általában nem foglalkoznak, mert a nyelv problémáit csak mint statisztikai–kvantitatív kérdést kezelik. A nyelv szempontjából ezeket az eljárásokat kvantitatív nyelvészeti eljárásoknak nevezik.

Mindezt az automatizált eljárásokról nem csak azért kellett elmondani, mert egyre jobban el fognak terjedni, még a könyvtári információkeresésen belül is. Elsősorban azért mondtuk el, hogy rávilágítsunk: **valójában a keresés teljesen soha sem formalizálható és ezért teljesen nem is gépesíthető.** Az információkeresés egy része mindig ún. hiperkomplex, az emberi intuíción alapuló, szubjektív eljárás marad, melynek vannak stratégiai, taktikai és szükségképpen lélektani összetevői is.

Azoknak az eljárásoknak, módszereknek az összességét, amelyekkel megszerkesztik a keresőképet és elvégzik az összehasonlítást, nevezzük **keresési stratégiának** és taktikának (az utóbbival a “3.7.2 Kereső taktikák” című fejezetben foglalkozunk részletesen).

A keresési stratégia alkalmazása ismétlődő és közelítő (iteratív) tevékenység, mivel a kapott találatok elemzése alapján a keresőprofil módosítására és az összehasonlítás megismétlésére kerülhet sor annak érdekében, hogy a tárolót optimálisan lekérdezhessük.

A keresési stratégia módszereire jelentősen hatottak a 60-as évekbeli cranfieldi vizsgálatokkal kezdődő hatékonysági kísérletek. Az első stratégiai összefoglalásokat Frederick W. Lancaster írta.⁷

A taktikai megközelítések még korábbra, a század harmincas éveire vezethetői vissza. Az ilyen jellegű kutatások legjelentősebb képviselője Marcia John Bates, aki a hetvenes években publikálta kísérleti eredményeit⁸ A javasolt heurisztikus stratégiai és taktikai módszerek többnyire más természetű esetekre alkalmazott logikai módszereken, tudományos ötleteken és tapasztalatokon alapulnak. A keresési stratégia legfontosabb fogalmait Charles Meadow, főleg pedig Frederick W. Lancaster dolgozták ki és vezették be a gyakorlatba.⁹

3.6.2 Online keresési stratégiák

A keresési stratégia fogalmát különféle értelemben használják az információkeresésben. Néhány szerző már a keresőkép megszerkesztését stratégiának nevezi, noha ezt helyesebb a keresőkérdés

⁷ A cranfieldi vizsgálatokra vonatkozóan lásd a Szöveggyűjtemény 1. kötetében az “Információkeresés értékelése” című részt, p. 195–202.

⁸ Részletesebben lásd a Szöveggyűjtemény “Heurisztikus és lélektani megközelítés, avagy a szubjektív tényezők megragadásának kísérlete” című részt, p. 340–365.

⁹ Lásd a Szöveggyűjtemény 1. kötetében a “Gépi információkeresés klasszikusai” (p. 169–190) és a 2. kötetben “Az online információkeresés elterjedése és a kézikönyvek” (p. 366–429) című részeket.

megformálásának tekinteni. Ebben a könyvben ez utóbbi felfogás szerint jártunk el, a keresőkérdés megformálását és a keresőkép szerkesztését a keresés első lépésének tekintjük, és eredményét használjuk föl mintegy “stratégiailag”.

Mind a stratégia, mind a taktika meghatározott szempontok érvényesülését jelenti a keresés folyamatán belül; annak eldöntését, hogy mi a keresés lebonyolításának legjobb útja. A stratégiában a már megformált kérdések konkrét, módszeres alkalmazásáról van szó, a keresés egyes útjainak pontos tervéről.

A taktika ennél intuitívabb módszerek összessége. Mind a stratégia, mind a taktika lényegében heurisztikus eljárás: a feltalálás, a rájövés módszerének tudománya, a rávezető, kitaláló módszerek összessége, mely ötleteken, tapasztalatokon és intuíción alapul.

A továbbiakban a legfontosabb stratégiai eljárásokat foglaljuk össze Stephen P. Harter, Charles Meadow és Richard Hartley nyomán. A felsorolt eljárások egymással kombinálhatók is.

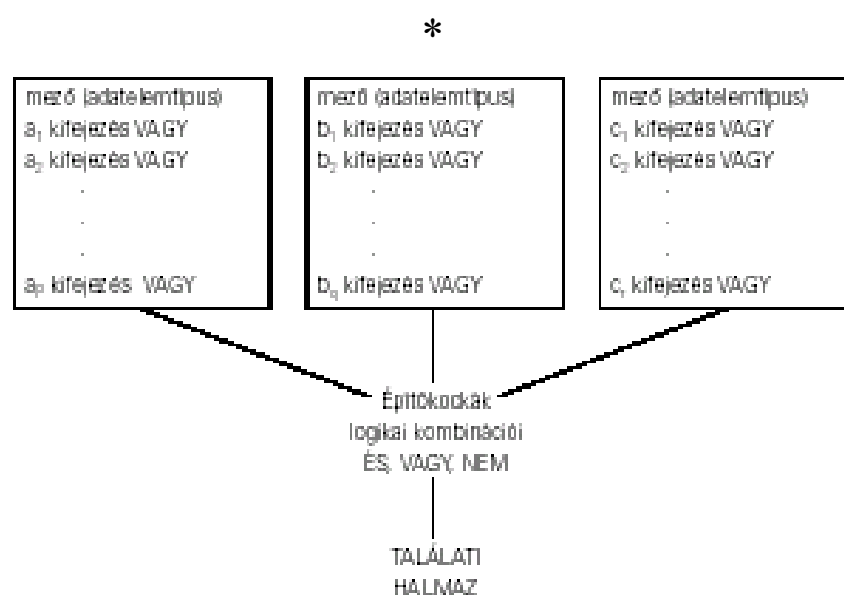
3.6.2.1 Egyszerű gyorskeresés

Ebben az eljárásban (angolul briefsearch, quick and dirty search, vagyis “rövid keresés”, “gyors és zajos keresés”) Boole-operátorokat alkalmazva rövidebbre szabott, gyors keresésre kerül sor annak érdekében, hogy előzetes kép alakuljon ki az adatbázis rekordjainak a kérdéssel kapcsolatos jellemzőiről. A felhasználó által közvetlenül megadott kifejezésekkel vagy azok csonkolt változataival hajtjuk végre. A kapott találati tételekből tapasztalatokat szűrhetünk le a részletes keresés – pontosabb keresőkép – megvalósításához.

3.6.2.2 Fogalmi építőkockák alkotása

A fogalmi építőkockák alkotásán (building blocks) alapuló stratégiát akkor célszerű alkalmazni, ha a keresőkérdés több szóból áll; egy tárgy részletező leírásáról van szó.

1. Megfogalmazzuk a keresőkérdést.
2. A kérdésen belül minden keresőszó egy-egy keresőfogalmat nevez meg. Minden fogalomnak egy-egy építőkocka a kerete, amelyen belül valójában nem mindig csak egy, hanem olykor több azonos vagy közel azonos jelentésű keresőszóval lehet megnevezni az adott fogalmat. A keresőszavak alapján VAGY műveleti jelet használva építőkockaként keresőképeket kell szerkeszteni.
3. Az egyes építőkockákat összekapcsoljuk értelemszerűen megválasztott Boole-operátorokkal. Ez lesz a teljes keresőkép.
4. Az eredményt elemezzük, szükség szerint visszacsatolunk és módosítunk a keresőképen.



15. ábra. A keresőkérdés építőkockái

*

Az egyes keresőfogalmakat az építőkockába vagylagosan összevont kifejezések együttesen nevezik meg.

A legfontosabb, hogy minden azonos vagy közel azonos jelentésű, de akár átfogóbb vagy speciálisabb keresőszót megadjunk. Az építőkockát tehát **az adott keresés szempontjából ekvivalens** ismérvek alkotják (az építőkockán belül ugyan lehetnek speciálisabb és általánosabb fogalmak kifejezései is, a szerepük itt azonos: a keresőfogalom egy-egy elemét képviselik). A teljes keresőkérdésben tetszés szerinti számú építőkocka is lehet.

Például: “ásványolajos szennyeződések meghatározása a talajvízben”.

F1 ásványolaj? VAGY kőolaj? VAGY olaj? VAGY ((Diesel VAGY Dízel EGYÜTT olaj)

F2 szenny? VAGY talajszenny?

F3 meghatározás VAGY azonosítás VAGY elemzés VAGY analí? VAGY anali?

F4 talajvíz Vagy talaj VAGY föld

F5 F1 ÉS F2 ÉS F3 ÉS F4

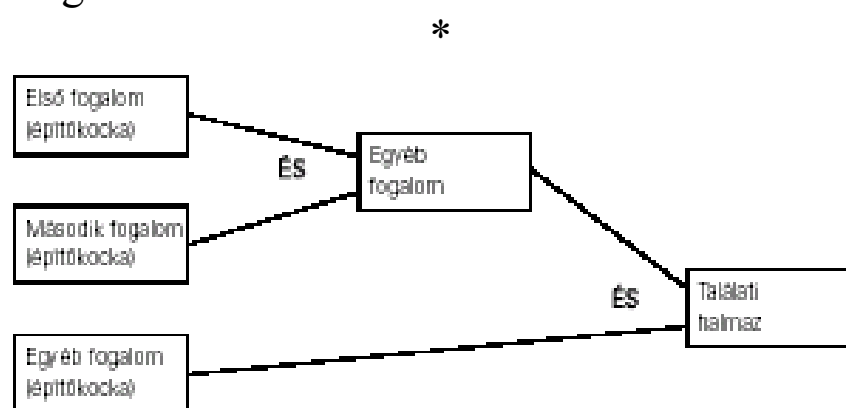
A ? a jobb oldali csonkolás jele. A “szenny?” esetében például keresőszónak számítanak a következők: “szenny”, “szennyeződés”, “szennyeződések”, “szennyforrás”, “szennyezőanyagok” stb.

3.6.2.3 Keresőszavak egymás utáni leválasztása

A teljesség növelésének heurisztikus módszere a keresőszavak egymás utáni leválasztása (angol neve successive facet strategies). Ilyenkor valamelyik kész építőkockát (keresőfogalmat reprezentáló kifejezések csoportját) vagy annak/azoknak egyes alkotóit töröljük és a maradékkal keresünk. Fokozatosan leválasztjuk a felesleges találatokat egészen addig, míg megfelelő méretű releváns találati halmazt nem kapunk. Ez persze a leggazdaságtalanabb eljárás, hiszen gondosan megszerkesztett építőkockáról mondunk le.

Ha feltehető, hogy az összes fontos keresőfogalom (építőkocka) felhasználásával túl kevés a találat (vagy egyáltalán nincs találat), vagy egyes építőkockák csak bizonytalanul reprezentálnak valamilyen keresőfogalmat, akkor módosítani kell a keresőfogalmakon alapuló stratégiát. A módosítás lényege, hogy a nagyobb teljességet (sok találatot) eredményező kérdésformával kezdjük a keresést, és a következő kérdésformák fokozatosan csökkentik a találati halmazt a már kezelhető, értelmes méretre.

Az eljárást az alábbi ábrán foglaltuk össze:



16. ábra. Az egymás utáni leválogatás módszere

*

Az eljárás alapelve, hogy az első két keresőfogalomból ÉS–kapcsolattal származtatott keresőfogalmakat (származtatott építőkockákat) egyesével kell megszerkesztteni, lépésről-lépésre és mindig szükség szerint, nem pedig mindet egyszerre. Minden lépésben az előző keresési eredmény segítségével szerkesztik meg a következő új származtatott keresőfogalmat. A keresés akkor ér véget, amikor a teljesség és a pontosság kívánt mértékét sikerült elérni, előnyös esetben még azelőtt, hogy az összes szóba jöhető építőkockát meg kellett volna szerkesztteni.

Az egymás utáni leválogatás tehát úgy kezdődik, hogy első lépésben létrehozunk egy kezdő találati halmazt, és annak méretét szükség szerint, lépésről-lépésre csökkentjük. A kezdő találati halmaz

gyakran dokumentumtípus, nyelv, vagy a kiadás éve. Mindezek a stratégiák olyan találati halmazzal kezdődnek, melyre a nagy teljesség a jellemző, és ennek mértékét csökkentik a további lépésekben az ÉS operátorral összekapcsolt fogalmakkal vagy más módon. Különösen olyankor célszerű a lépésenkénti leválogatás stratégiájával élni, amikor a keresés tárgya, témája nem körvonalazható pontosan, vagy túl általános, vagy egyéb okból hasznos, de nem lényeges korlátozások engedhetők meg a kérdésben.

1. példa: “az Ördögűző Klub tagjai és a klub tevékenysége”
A kérdésnek megfelelő keresőkép Boole-operátorokkal ördögűzők klubja ÉS (tagok VAGY tevékenység.) Nyomós okok szólhatnak amellett, hogy másként célszerű megközelíteni a megoldást.

Egyrészt: a “tevékenység” nagyon általános fogalmat nevez meg. Klubtagok tevékenysége az égvilágon mindenféle lehet, de esetünkben nem “ördögűzés”, mivel a szóban forgó klubnak ez amolyan “blikkfangos” neve.

Másrészt: az “Ördögűző Klub” rendkívül specifikus kifejezés, valószínűtlen, hogy van róluk információ adatbázisokban. Harmadrészt: a keresőkérdésbe fölvetett keresőszó az feltételezi, hogy a szó ismérvként a dokumentumképben is megvan. A “tag” vagy a “tevékenység” keresőszavaknak is csak akkor van értelme, ha dokumentumokat is osztályoznak e két kifejezéssel, melyek meglehetősen tág értelmet adnak a dokumentum tartalmának. Valószínű, hogy sok releváns dokumentumrekord nem tartalmazza a “tevékenység” kifejezést, ellenben tartalmaz a lehetséges speciális tevékenységyszavak közül néhányat.

Végül azzal lehet még érvelni, hogy azok a dokumentumok, melyek az “Ördögűző Klubról” szólnak, esetleg tagokról és a tevékenységekről is szólnak. De persze szólhatnak a szervezeti felépítésről, és akkor a tagokról meg a tevékenységről nem sokat tudhatunk meg, de valami keveset azért mégis.

Első lépésben tehát csak az “Ördögűző Klub” kifejezéssel keressünk. Ha elég releváns találatot kaptunk, és ezek elég tartalmasnak látszanak, fejezzük be a keresést.

Ha a találati halmaz túl nagy vagy túl kicsi, vagy a pontossága elfogadhatatlan, folytatható a keresés.

2. példa: Oktatási/nevelési adatbázisban a “Kodály-módszer alkalmazása a zeneoktatásban” kérdés esetén elég a “Kodály” nevet megadni, hogy releváns dokumentumokat kapjunk, mert Kodály zenei munkássága nem tartozik az adatbázis gyűjtőkörébe; zenei adatbázisban viszont az “oktatás” keresőszóval célszerű elkezdni a keresést, és ennek eredményétől kezdve szűkíthetünk.

3. példa: “Diszlexiás gyerekek, ill. gyermekek írás-, olvasásoktatása az alapfokú oktatási intézményekben”.

Attól tarthatunk, hogy ha a szokványos keresőképet adjuk meg: ((diszlexi? ÉS gyer?k?) ÉS (írás VAGY olvasás)) ÉS (alapfokú oktatási intézmény? VAGY alapfokú oktatás), nem sok találatot kapunk, ha egyáltalán kapunk. Ezért először lekérdezzük a “diszlexia” illetve “diszlexiás” keresőszavakkal (a “gyer?k” szót elve elhagyjuk), ha túl sok a találat vagy nem elég pontos, finomítunk az “írás” és az “olvasás” felhasználásával, és csak ha még ezek után érdemes finomítani a találati halmazzal, használjuk föl az “oktatás” vagy az “alapfokú oktatás” vagy az “alapfokú oktatási intézmény” keresőszavakat.

F1 diszlexi? ha nem elég:

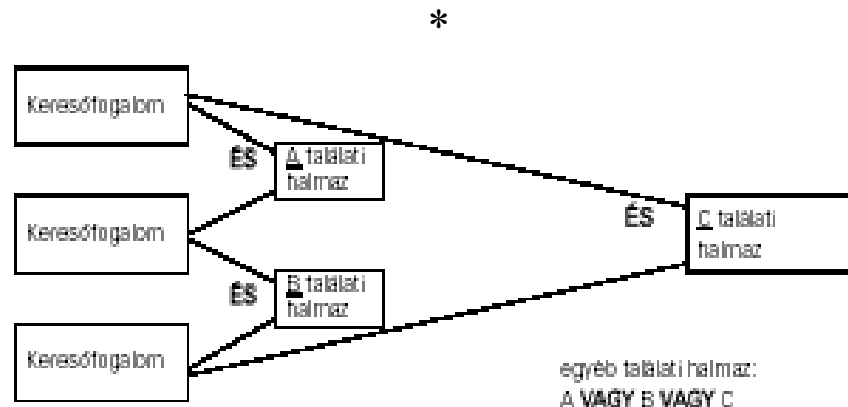
F2 F1 ÉS (írás VAGY olvasás) ha nem elég:

F3 F2 ÉS ((elemi VAGY általános) EGYÜTT (alapfokú oktatás) VAGY oktatási intézmény

F4 eredményhalmaz

3.6.2.4 Páronkénti leválogatás

Ha mindegyik keresőfogalom (építőkocka) nagyjából azonos mértékben specifikus, azaz egyformán fontos, akkor páronként képezhetjük metszetüket. A találati eredményeket vagy külön-külön (a páronkénti keresés eredményeként), vagy az egész keresési folyamat uniójaként nyomtathatjuk ki. A stratégiát (angol neve pairwise facets) az alábbi ábra mutatja be:



17. ábra. A páronkénti leválogatás

*

Például: “arcizmok, a fiziológiai reakciók és a félelem összefüggése”. A doktorandusz abban érdekelt, hogy mennél több hivatkozást sorolhasson föl a bibliográfiájában. Számára tehát releváns lehet az arcizmok és a fiziológiai reakciók összefüggése, a fiziológiai reakciók és a félelem összefüggése stb. Az “F0: (arcizom VAGY arcizmok) ÉS fiziológiai reakció? ÉS (félelem VAGY rettegés)” eredménye nem volt kielégítő, mert csak nagyon kevés találatot eredményezett.

F1	(arcizom VAGY arcizmok) ÉS fiziológiai reakció?	A találati halmaz
F2	(arcizom VAGY arcizmok) ÉS (félelem VAGY rettegés)	B találati halmaz
F3	fiziológiai reakció? ÉS (félelem VAGY rettegés)	C találati halmaz
F4	F1 VAGY F2 VAGY F3	egyéb találati halmaz

3.6.2.5 Többszörös egyszerű gyorskeresés

Ezt a stratégiát (angol neve multiple briefsearch) a páronkénti leválogatás esetéhez hasonlóan akkor célszerű követni, ha az összes specifikus keresőfogalom metszetének eredménye várhatóan a nullához közelít. Alapelve, hogy – lehetőleg különböző szolgáltatóknál – **több adatbázisban** kell egyszerű, nagy teljességgel járó gyorskeresést végezni. Ezáltal nagyon különböző jellemzőket kapunk, ami a keresés témáját illeti, melyet a további lépésekben hasznosíthatunk. Ugyanazt a tárgykört az egyes adatbázisokban ugyanis más-más módon osztályozzák (indexelik) és dolgozzák föl formai szempontból. Ezáltal sok olyan jellemzőt megállapíthatunk, melyeket egyetlen adatbázisból körülményesebb kielemezni.

3.6.2.6 Hólabdakeresés

Ezt a stratégiát (angol neve pearl growing) olyankor célszerű használni, ha nagyon kevés találatot várunk. Ellentétben az eddigi stratégiákkal, itt nem a nagy teljességgel kezdünk, hanem éppen ellenkezőleg. A leggyakoribb, hogy a felhasználó megadja az egyik, általa ismert és a tárgyba vágó dokumentum bibliográfiai adatát, s ennek alapján megkeressük a dokumentumtételt. Ebből megállapítható, milyen ismérvek (deskriptorok, tárgyszavak, jelzetek) reprezentálják a tartalmát –

tehát a keresett tárgy fogalmát. Most már ezekkel az ismérvekkel végezhető a keresés, hogy még több hasonló dokumentumtételhez jussunk stb.

- A keresőszó = ismert releváns "hólabda-rekord"
B keresőszó ÉS C keresőszó = keresési eredmény ("hólabda-rekordban" talált kifejezésekkel)

Például: szeretnénk ETO-jelzet szerint is keresni az "online és CD-ROM adatbázisokban végzett információkeresés" témáját, de nem ismerjük a megfelelő ETO-jelzete(ke)t. Azt tudjuk, hogy a tárgykörben Roboz Péter publikált. Keresünk tehát a "Roboz Péter" személynév szerint, és megállapítjuk, hogy a kapott két találathoz az alábbi ETO-jelzetek kapcsolódnak (megadjuk a jelzetek magyarázatát is, melyek a dokumentumképben természetesen nem szerepelnek):

1. találat	2. találat
025.5 Felvilágosítás. Referenz-szolgálat	025.4 Információkereső nyelvek, osztályozási rendszerek
659.2 Információ, tájékoztatás, tanácsadás	061.68 Adatközpontok, adatbankok
681.3.016 File-szervezés és -feldolgozás	681.3.015 Párbeszéd (interaktív) üzemmód
681.3.004.14 Számítógépek alkalmazása	681.324.004.14 Automatikus adatfeldolgozás alkalmazása

A két találatból álló eredmény elégtelen. A keresést megismételjük a kapott ETO-jelzetekkel. Már lényegesen több találatot kapunk. Az ezekhez kapcsolt ETO-jelzeteket is elemezve alakíthatjuk ki azt az ETO-jelzetekből álló keresőképet, melyet optimálisnak tartunk.

3.6.2.7 Relevancia-visszacsatolás

A relevancia-visszacsatolás (relevance feedback) emlékeztet a hólabdakeresésre. Ezt a stratégiát akkor alkalmazzuk, amikor van ugyan elég találatunk, de pontosítani szeretnénk a keresést. Lépései:

- valamilyen keresési stratégia alapján elvégzünk egy vagy több keresést;
- kiválogatjuk a legrelevánsabb első n számú találatot;
- ezeknek a találatoknak az ismérveivel újabb keresést hajtunk végre.

A parancsnyelvekben erre a stratégiára többnyire rendelkezésre áll parancs (a DIALOG rendszerben RANK a neve). A stratégia konkrét alkalmazási példája látható a 14. ábrán.

3.6.2.8 Összefoglalás

Az alábbiakban foglaltuk össze az egyes stratégiákat (az ábécé A, B, C nagybetűi, illetve a nagy- és kisbetű- kombinációk – mint Aa, Ab stb. – a keresőszavakat jelölik, ha nincsenek találati halmazként megnevezve):

Egyszerű gyorskeresés

A ÉS B = Találati halmaz

Fogalmi építőkövek alkotása

Aa VAGY Ab VAGY Ac = 1. halmaz (nagy halmaz)

Ba VAGY Bb VAGY Bc = 2. halmaz (nagy halmaz)

Ca VAGY Cb VAGY Cc = 3. halmaz (nagy halmaz)

1. halmaz ÉS 2. halmaz ÉS 3. halmaz = Találati halmaz

Keresőszavak egymás utáni leválogatás

- A ÉS B = 1. halmaz (nagy halmaz)
1. halmaz ÉS C = 1. halmaz származtatott része
2. halmaz ÉS D = Találati halmaz

Páronkénti leválogatás

1. FOGALOM ÉS 2. FOGALOM = B találati halmaz
2. FOGALOM ÉS 3. FOGALOM = C találati halmaz
3. FOGALOM ÉS 1. FOGALOM = A találati halmaz
A találati halmaz VAGY B találati halmaz VAGY C találati halmaz = egyéb találati halmaz

Többszörös egyszerű gyorskeresés

1. adatbázisban: A ÉS B = 1. találati halmaz
2. adatbázisban: A ÉS B = 2. találati halmaz
3. adatbázisban: A ÉS B = 3. találati halmaz

Hólabdakeresés

- Keresés bibliográfiai adatok alapján = 1. találati halmaz (1 vagy néhány találat)

A további kereséshez a találatokat jellemző tartalmi ismérvek alapján állapítjuk meg a keresőszavakat

- Keresés az 1. halmaz találatának/találatainak tartalmi ismérvei alapján = 2. találati halmaz (több találat)

- Keresés a 2. halmaz találatainak tartalmi ismérvei alapján = 3. találati halmaz (sok találat)

Relevancia-visszacsatolás

1. keresés = 1. találati halmaz
2. keresés megismétlése az 1. találati halmaz legrelevánsabb n találatának ismérveiből szerkesztett keresőképpel stb. = 2., nagyobb relevanciájú találati halmaz

3.7 Az információkeresést megkönnyítő taktikák

3.7.1 A taktikák szerepe

Az információkeresés automatizálásában elért eredmények ellenére semmi nem ér még fel a tapasztalt ember ilyen irányú tudásával, képességeivel. Ugyanakkor keveset tudunk ezekről az ismeretekről, és azt sem tudjuk, mi is voltaképpen a különbség e téren a tapasztalt szakember és a kezdő tudása között.

A keresést megkönnyítő taktikák bibliográfiai és faktografikus (tényadatokra vonatkozó) keresésekre, a legkülönbözőbb keresési típusokra és témakörökre, manuális és online rendszerekre egyaránt alkalmazhatók. A taktikákat elsősorban komplexebb, többlépcsős keresésekhez alkalmazzák, illetve azokra a szituációkra, amelyekben még semmit sem tudni a lehetséges forrásokról. Néhány taktika közismert, de még ezek közül sem volt korábban mindegyiknek neve, és vannak köztük kevésbé ismertek is. Csoportosításukkal és megnevezésükkel először Marcia John Bates foglalkozott.¹⁰ Egy-egy taktika világos leírása és elnevezése már önmagában is segítség: tudatosabbá teszi alkalmazását, és gyorsabban “ugrik be” a kereső emlékezetébe.

¹⁰ Szöveggyűjtemény, 2. kötet, p. 347–358.

A taktikák egész halmazát tekintve vannak közöttük átfedések és bizonyos hierarchikus viszonyok is. Ezeknek a taktikáknak a használata a kreatív problémamegoldás egyik formája. Ilyen esetekben a gondolkodás nem mindig logikus, szabályos sémák szerint működik. Sok különböző irányból közelíthetünk meg egy-egy problémát, egyszer ilyen, másszor pedig olyan taktikát alkalmazhatunk ugyanabban a problematikus esetben. Más szóval az alábbi taktikák között szinte követelmény a bizonyos fokú redundancia.

A keresési taktikák elsősorban a keresés egészére vonatkoznak, nem pedig az egyes lépések konkrét megfogalmazására (az utóbbi a stratégia feladata). Elsősorban rávezető szerepük van, fortélyoknak is nevezhetők. A keresőkép szerkesztéshez és a keresési stratégiához képest a keresési taktika a leginkább heurisztikus módszer.

A felsorolt taktikák valószínűleg javítják a keresések hatékonyságát. A valószínűleg szó azonban fontos, mivel ezek a taktikák heurisztikusak, s ezért nem szükségszerű, hogy segítsenek. Továbbá, valamely taktika jó lehet az egyik szituációban, de egy másikban nem.

A hatékonyságot közvetlenül javító taktikákat kereső taktikáknak, a közvetve javító taktikákat pedig képzelettaktikáknak nevezzük.

A kereső taktikáknak négy fajtája van:

1. Felügyelő taktikák. Céljuk, hogy a keresést a megfelelő vágányon tartsák, és a hatékonyságra figyeljenek.
2. Fájlszerkezet-taktikák. Hogyan találja meg a kereső az utat a széles értelemben vett információhordozók között, a kívánt forrásig vagy a forrásban lévő információig?
3. A kérdés megfogalmazásával kapcsolatos taktikák. A keresőkép tervezésében, módosításában segítenek: nem korlátozódnak a számítógépes keresésekre.
4. Szóhasználati taktikák. A kérdés megfogalmazása közben a kifejezések kiválasztásának módosításában segítenek.

A taktikák tárgyalásakor (mivel nemcsak online kereséskor alkalmazhatók) a fájl (a rendezett állomány) fogalmát a megszokottnál tágabban értelmezzük: beletartozik az információs egyedek bármely rendezett halmaza. A könyv tipikusan ilyen információs egyed: a könyvtár bibliográfiai állománya szintén fájl.

A fájlszerkezet jelentése értelemszerűen szintén tágul. Például minden indexelő- rendszerhez tartozik valamilyen szerkezet, ám most nem ennek milyensége, hanem meglétének pusztá ténye a fontos.

A keresési taktikák alább ismertetett fajtáinak megfogalmazója, Bates végcélként fölvetette, hogy kidolgozható volna egy átfogó taktikakészlet. Lehetővé válna általa, hogy egységes szemlélettel tekintsük át az egész referenz folyamatot, és magját képezhetné egy referenzkönyvtáros-képző és információkeresési tananyag.

3.7.2 Kereső taktikák

3.7.2.1 Felügyelő taktikák

ELLENŐRZÉS

Az eredeti kérdés összehasonlítása a keresés pillanatnyi tárgyával: azonosak-e még.

MÉRLEGELÉS

A keresés egy vagy több pontján felbecsüljük, mennyibe kerülnek vagy mit eredményeznek a következő lépések. Megfontolható, hogy másféle megközelítés nem volna-e eredményesebb.

KAPTÁFA

A gyakori típuskérdések megszokott keresési sémákhoz vezetnek. Itt arról van szó, hogy a kérdés tipikussága és az előhúzó kaptafa tudatosuljon a keresőben, a megoldási sémát azonban át kell gondolni, és ha nem maximálisan hatékony vagy esetleg már idejétmúlt, akkor módosítani kell.

JAVÍTÁS

Nemcsak az a fontos, hogy a keresés tárgyát pontosan meghatározzuk, hanem az is, hogy ezt a tárgyat a helyesírás szabályai szerint fogalmazzuk meg. A helyesírási hibát elkövetheti a feldolgozó is, de most elsősorban a keresőről van szó, amikor például emlékezetből keres, nem lévén az írás a kezében. Így fordul elő, hogy neurológia helyett neuralgia lesz a keresés tárgya, ami pedig igencsak eltérő fogalom.

JEGYZETELÉS

Maradjon nyoma, hogy milyen utakat járt végig a kereső, melyeket szakított meg, sőt annak is, hogy mit akart volna még megvalósítani, de nem került rá sor.

3.7.2.2 Fájlszerkezet-taktikák

“BIBLIZÉS”

Ez az elnevezés neologizmus, a bibliográfia rövidítéséből származik. Kész bibliográfia keresését jelenti, mielőtt nekiállnánk, hogy készítsünk egy újat. Általánosabban fogalmazva, annak ellenőrzéséről van szó, hogy a keresést nem végezték-e már el, az eredmény nem elérhető-e valamilyen használható formában.

PROBLÉMALEBONTÁS

Komplex kérések lebontása részproblémákra. Az egyes részeken külön, egymás után lehet dolgozni, és a részmegoldásokat össze lehet kötni az egész probléma megoldásává.

ÁTTEKINTÉS

A választható lehetőségek áttekintése keresések közben a döntések előtt. Helytelen idő előtt leragadni egyetlen forrás vagy megközelítési mód mellett. Az áttekintés révén az ember ellenállhat az ilyen kísértésnek. Például az elsőnek éppen eszünkbe jutó online mutató böngészése helyett gondolatban vegyük sorra a téma szempontjából releváns összes mutatót, és válasszuk ki közülük az adott kéréshez leginkább illőt, azután pedig ne kössünk ki rögtön egy mutatószónál, hanem kezdjük el böngészni az online mutatóban, onnan meg átlépve a teauruszban, keressük meg benne a legjobb keresőszóként használható kifejezést.

KIZÁRÁS

Nagyon fontos taktika. Ha többféleképpen kereshetünk, válasszuk azt a lehetőséget, amely azonnal a lehető legtöbb, vélelmezhetően zajos találatot zárja ki a további keresésből. Azaz, ha az operációs rendszerek parancsnyelvéről szóló könyveket, vagy a Kis János és Bencze György szerzőpáros filozófiai könyveit keressük, az első esetben a “parancsnyelv” (mivel operációs rendszerekről szóló dokumentumoknak se szeri, se száma), a második esetben az utóbbi név alapján lényegesen hamarabb jutunk eredményre.

KITERJESZTÉS

Az információforrásokra természetesen olyan összefüggésben szoktunk gondolni, amilyen jellegű használatra szántuk őket. Mégis, szinte minden forrás eredményesen használható nem tervezett célokra is. Ehhez persze nemcsak a rutinszerű használati módokat, hanem a forrás egész információtartalmát ismerni kell. Ha például sikertelen volt a nyomozás egy mérnök címe után, a keresőnek eszébe juthat, hogy a szabadalmakon a feltaláló neve mellett rendszerint a munkáltató is szerepel. Ezért a keresést a szabadalmak között végezzük el a mérnök neve szerint. Amennyiben az illető mérnöknek van valami szabadalma, a munkáltató címe is meglesz a szabadalmi nyilvántartásban, ennek alapján pedig már előbb-utóbb elérhető lesz a mérnök címe is.

KÖRÜLÁLLVÁNYOZÁS

Amikor az épület elkészült, az állványokat lebontják, de nélkülük az épület nem épült volna fel. Az információkeresés néha ugyancsak ilyen körülállványozásra kényszerül. Olyan

információelemekkel kell dolgoznia, amelyeknek közvetlenül ugyan semmi közük sincs a válaszhoz, de lehet, hogy végül mégiscsak hozzásegítenek. Ha például egy jelentéktelen költőről nem sikerül információt találni, a kereső utánanézhethet, kik voltak a kortársak, és náluk kereshet tovább, hátha valahol említik az illetőt.

3.7.2.3 A kérdés megfogalmazásával kapcsolatos taktikák

SPECIFIKUSSÁG

Minden osztályozási rendszerben és információkereső nyelvben megkövetelik, hogy amennyire csak lehet, a dokumentumok tartalmi leírásai a tartalomnak megfelelő mértékben specifikusak legyenek. Előfordulhat ugyan, hogy a dokumentumtételleket átfogóbb fogalmak alá is be kell sorolni, de ebben az esetben is fontos, hogy a specifikus fogalmak alatt is szerepeljenek. Lehet például úgy eljárni a feldolgozáskor, hogy a márkanévek esetén az átfogó fogalommal is mindig osztályozzanak, de a márkanéveket ismervként ebben az esetben is hozzá kell kapcsolni a dokumentumtételhez (Például a Ford Escort autómárkáról szóló dokumentumot osztályozzák a “személygépkocsi” deskriptorral, de “szabad tárgyszóként” a “Ford Escort” márkanévet is a dokumentumtételhez kapcsolják.) Ezért a keresést is a specifikus terminusokkal érdemes kezdeni.

RÉSZLETES KÉRDÉSMEGFOGALMAZÁS

Ez a taktika a kérdés első megfogalmazásába beviszi annak összes vagy legtöbb elemét, vagy a már kész keresőképet bővíti egy vagy több keresési elemmel. Minél bővebb, azaz minél több elemet kapcsolnak össze ÉS relációval a keresőképben, annál szűkebb a kérdések tartalma, jelentése, tehát annál kevesebb találat várható eredményképpen.

SZÜKSZAVÚ KÉRDÉS-MEGFOGALMAZÁS

Ez a taktika a kérdés első megfogalmazásában minimalizálja az elemek számát, vagy a már kész keresőképből vesz el egy vagy több elemet. Minél kevesebb elemet kapcsolnak össze ÉS relációval a keresőképben, annál tágabb a kérdés, tehát annál több találat várható eredményképpen.

HASONLÓSÁGOK MEGENGEDÉSE

A keresőkép szélesítése, bővítése szinonimák vagy egyéb – rokonértelmű – kifejezések bevonása által; voltaképpen a VAGY kapcsolat bőséges alkalmazásáról van szó.

SZÜKSZAVÚBB MEGFOGALMAZÁS

Az előző taktika ellentéte, a keresőkép lehető legpontosabb megfogalmazása a rokonértelmű kifejezések számának minimalizálása vagy legalábbis csökkentése és a legtalálhatóbb kifejezések megtartása által.

KIZÁRÁS

A kérdés megfogalmazásával kizárjuk a válaszból azokat a keresési szempontokat, amelyek – önmaguk vagy mutatóik – bizonyos kifejezés(ek)e)t tartalmaznak, azon az áron is, hogy releváns dokumentumokat veszítünk szem elől. Ez a taktika voltaképpen a DE NEM logikai művelet megfelelője. Azért kizárás a neve, hogy a NEM fent említett kényes oldalára felhívja a figyelmet. A nemkívánatos kifejezést tartalmazó dokumentumok kizárása kívánatos információ elvesztésével járhat.

3.7.2.4 Fogalom- és szóhasználati taktikák

FELJEBB LÉPÉS (Hierarchiaszint-emelés)

Feljebb lépés a hierarchiában az általánosabb, fölérendelt kifejezéshez. A keresőt segítheti a teaurusz vagy – különösen – az osztályozási rendszer, de lehet, hogy saját ismereteire támaszkodva kell ezt a kifejezést megállapítania.

LEJEBB LÉPÉS (Hierarchiaszint-csökkentés)

Lejjebb lépés a hierarchiában egy specifikusabb, alárendelt fogalomhoz.

ASSZOCIÁCIÓ (Bővítés rokonsági kapcsolatok bevonásával)

Rokonértelmű vagy a keresett kifejezéssel gyakran együtt előforduló, illetve a keresett kifejezésről eszünkbe jutó kifejezés keresése – “oldalirányú” lépés a hierarchiaszinten belül.

SZOMSZÉDKERESÉS

Ez a taktika a szomszédos kifejezések között keres továbbiakat, akár a betűrend, akár a tartalmi hasonlatosság, akár valami más szomszédság alapján. A szomszédkeresés az adatbázisok vagy a kézikönyvek kiválasztásra is kiterjeszhető, például amikor megvizsgáljuk a referenzszolgálat polcain egymás mellett tartott kézikönyveket.

NYOMKÖVETÉS

A már fellelt információ megvizsgálása olyan újabb kifejezésekért, amelyek továbbvihetik a keresést. Egyik mindennapi formája az online kereséskor kapott találatok dokumentumképének áttekintése. Végignézzük az ismérveket, amelyek az adott dokumentumtételben szerepelnek az éppen aktuális ismérven, keresőszón kívül. Ezeket az ismérveket “nyomoknak” hívják, innen származik a taktika elnevezése.

VARIÁCIÓK

A kifejezések módosítása, helyettesítése.

AFFIXUMOK VARIÁLÁSA

Próbálkozás különféle -nyelvi előtagokkal (prefixumokkal, például hajtás, *meghajtás*, *áthajtás*), utótagokkal (szuffixumokkal, például szenny, *szennyeződés*, *szennyezés*; viszály, *viszálykodás*) és köztes tagokkal (infixumokkal, például Abbe-elv, *Abbe-féle elv*), illetve elhagyásukkal. A csonkoló parancsokkal egyidejűleg több ilyen művelet is elvégezhető.

E nyelvi pre- és szuffixumok nem azonosak a DIALOG-ban a keresőképben megadható szerzői (AU), cím (TI) stb. szótárakat (mutatókat) azonosító hasonnevű parancsszavakkal (pl. “S Cable?AU, ti” keresés a “cable” kezdetű szavak szerint a szerzői és cím mutatóban).

SZÓRENDVÁLTOZTATÁS

Minden olyan rendszerben, ahol egy kifejezés több szóból is állhat, a szórend befolyásolhatja a keresés sikerét. A többszavas terminusok esetén minden lehetséges – vagy legalábbis többféle – értelmes szórendet érdemes kipróbálni.

ELLENTÉT SZERINTI KERESÉS

A kívánt információt leíró kifejezés logikai ellentétével keresünk, például a “színes film” sikertelensége esetén a “fekete-fehér film”, a “koldulás” sikertelensége esetén az “adakozás” vagy az “alamizna” terminussal próbálkozunk.

HELYESÍRÁSI VÁLTOZATOK SZERINTI KERESÉS

A felügyelő taktikák egyike, a JAVÍTÁS egyebek között a jó helyesírásra hívja föl a figyelmet. A jelen taktikánál nem a helyességen, hanem a hatékonyságon van a hangsúly. Az online rendszerekben, még inkább az interneten nagyon tarka helyesírással találkozunk, s a jó eredmény érdekében gondolni kell a lehetséges változatokra (Diesel-, Dízel-, Dizel-, diesel-, dizel-, dízel-motor; Newton-féle törvény, Newton-törvény). Erre a taktikára az idegen nyelvű rendszerekben is szükség van, gondoljunk például csak a brit-amerikai (cataloguing, cataloging) vagy a német (telefon, telephon) helyesírási különbségekre. Különösen fontos az egybe- és különírási változatok figyelembe vétele (adatcsereformátum, adat-csereformátum, adatcsere formátum. Gondosan tanulmányozzuk a mutatókat!

A keresési taktikákkal kapcsolatban elég sok mindent meg kell fontolni, mielőtt alkalmazzuk őket.

1. Bizonyos taktikák csoportokat képeznek. Például a lehetséges reakciók olyan szituációkban, ahol a keresés túl sok vagy kevés találatot eredményez: a LEJEBB LÉPÉS, a RÉSZLETES KÉRDÉSMEGFOGALMAZÁS és a KIZÁRÁS, illetve a FELJEBB LÉPÉS, az ASSZOCIÁCIÓ, a SZÜKSZAVÚ KÉRDÉSMEGFOGALMAZÁS, a HASONLÓSÁGOK MEGENGEDÉSE, a NYOMKÖVETÉS és a VARIÁCIÓK.

Ha megkülönböztetjük a keresések tipikus szakaszait és megkeressük a nekik megfelelő taktikákat, kialakíthatunk stratégiai modelleket is. Ha a kereső tudja, hogy egy adott

szakaszban a taktikák melyik kis csoportjával számíthat leginkább sikerre, akkor abban a szakaszban csak erre a néhányra kell koncentrálnia.

2. A tényleges keresésen kívül a tájékoztatási folyamat más elemeire is kidolgozhatók taktikák. Ilyen területek a holtpontra jutott kérések kimozdítása, a referenzs-interjú, ennek részeként a konzultáció a felhasználóval a keresés előtt, alatt és után, a kérés kezdeti elemzésével kapcsolatos taktikák (például a számításba jövő források rendszere aszerint, mennyire valószínű, hogy ténylegesen segítenek, a felhasználótól jövő visszajelzéssel, a visszacsatolással kapcsolatos taktikák, és végül azok, amelyek az eredmény relevanciájának értékelésében segíthetnek). Ezek az eljárások már átvezetnek a képzelettaktikákhoz.
3. A könyvtári-informatikai gazdaságossági elemzések általában terjedelmes tanulmányokon és matematikai modelleken alapulnak. A MÉRLEGELÉS azzal foglalkozik, amit az emberek fejben, néhány másodperc alatt kiszámíthatnak. Egyszerű szabályokra van szükség, megalkotásuk azonban bonyolult munkát kíván. Míg a rendszerkutatók a gazdaságossági elemzések jól fejlett tudományát hívják segítségül, addig az információkeresők számára nincs ilyen tudomány, márpedig olyan keresési-döntési szabályok kellenek, amelyek minimalizálják a szellemi erőfeszítést.
4. A keresési stratégia egyik fő kérdése, hogy mikor kell megállni, hogyan állapítható meg, elég információ gyűlt-e össze, illetőleg mikor kell döntenünk a sikertelen keresés feladása mellett.

Az ÁTTEKINTÉSRŐL feltehető, hogy a keresés hatékonyságát minőségileg és mennyiségileg egyaránt javítja, de valahol határnak kell lennie; egy ponton túl egyre kisebb a haszon, csak hogy ezt a pontot még meg kell találni.

A MÉRLEGELÉSnek is vannak határai. Például miután a forrásokat relevanciájuk valószínűsége szerint sorba rendeztük, hol az az optimális pont, ahol az egyik forrásban való keresést abba kell hagyni, és át kell térni a következő forrásra? Alighanem jóval hamarabb érünk ehhez a ponthoz, mint hogy kimerítenénk az adott forrásban rejlő valamennyi lehetőséget.

3.7.3 Képzelettaktikák

A keresőtaktikák fizikailag létező, konkrét eszközök alkalmazására vonatkoznak; céljuk a keresés eredményességének közvetlen növelése.

A képzelettaktikák kizárólag értelmi műveletek, illetve ilyenekre irányuló tanácsok; céljuk, hogy a keresés holtpontjain átsegítsék a keresőt. A problémamegoldás közben gyakran elakad a gondolkodás, ilyenkor a kérdést új szemszögből, más módszerrel célszerű megközelíteni. Arra már ritkábban gondolunk, hogy az új irányba terelő ötleteknek gyakran az először alkalmazott, többnyire megszokott elképzelések állják útját. A képzelettaktikák új ötletek, megoldási módszerek, elképzelések keletkezését segítik elő, fellazítva a szellemi megkötöttségeket, olykor arra a felismerésre építve, hogy az információk szellemi feldolgozása és az adatok fizikai elhelyezése egymástól elválaszthatatlanok.

Az alábbiakban röviden közöljük az ismertté vált képzelettaktikákat. Ha azt akarjuk, hogy a taktikák eleven, készítő szerepet játsszanak bennünk, felszólító módba tett, tömör és kifejező igék formájában célszerű őket nyilvántartani.

E taktikák a képzelet fejlesztését, a bevált sémák feloldását és a képzeletben, illetve fizikailag tárolt információk egyeztetését segítik elő.

GONDOLKODJ!

A gondolkodás annyira magától értetődik, hogy meglehetősen pazarlóan bánunk vele. A kereséskor is sok időt és energiát fordítunk rá – sikertelenül. Úgy is mondhatjuk, hogy “gondolkozzatunk” ahelyett, hogy “gondolkodnánk”. A programozásban a GONDOLKODJ a jelszó szerepét játssza: azt jelenti, hogy “jusson eszünkbe valami helyes”, azaz vegyük

észre, hogy gondolkodni kell. Idézzünk elő és tartsunk fenn olyan szellemi állapotot, amelyben valóban gondolkodunk.

SZÁRNYALJ SZABADON! (ötletroham, brainstorm)

A kritikus értékelést teljesen félretéve engedjük utat szabad ötleteinknek, anélkül, hogy nyomban szabatosan meg is fogalmazzunk és értelmeznénk őket. Az ötletroham több ember aktív együttműködésén alapul, melyben kölcsönösen inspirálták egymást a szabad ötleteikkel, de – talán kisebb hatáskörrel ugyan – egyedül is szárnyalhatunk a gondolatainkkal.

MEDITÁLJ!

Minden felidéző, kereső folyamat tartalmaz befelé forduló, “képszerű” szakaszokat, melyek nehezen ragadhatók meg fogalmilag. E szakaszok aktivizálása érdekében nem annyira konkrét gondolatok szükségesek, mint inkább szellemi állapot, melyben az intuitív és a racionális gondolkodás párhuzamosan működik és egymásra talál. Ez az állapot a meditáció.

KONZULTÁLJ!

Ötletszerző, tájékozódó beszélgetés valaki mással a kérdésről.

MENTSŐ, AMI MÉG MENTHETŐ!

Vegyük elő a már eredménytelennek bizonyult megközelítéseinket és ellenőrizzük – nehogy a fürdővízzel együtt a gyereket is kiöntsük –, hogy van-e olyan változatuk, melyet még nem próbáltunk ki, mielőtt végleg más irányba indulnánk.

BÖNGÉSSZ!

Böngésszünk a segédletek és a szemünkbe ötlő egyéb eszközök között, hátha valamelyik alkalmas forrásnak bizonyul arra, hogy a keresés új kiindulópontja legyen.

OCSÚDJ! (tettenérés)

Vegyük észre, ha zsákutcába kerültünk, és jusson eszünkbe, hogy a megközelítés módosítása haszonnal járhat.

SÖPÖRD LE A MEGSZOKOTTAT!

Bizonyos fajtájú kérdések gyakrabban fordulnak elő, és megválaszolásuk is gyakrabban sikerül. Az ezekből szerzett tapasztalatok akarva-akaratlanul is a keresés általánosított formájának beidegzéséhez vezetnek. Ezzel az általánosított modellel viszonylag hamar és kevés fejtöréssel célhoz érhetünk. Előfordul azonban, hogy éppen ez okozza adott esetben a keresés sikertelenségét. Fordítsunk tehát hátat az eddigieknek, és söpörjük le a színről a megszokott keresési modellt (vagy legalábbis felejtsük el egy időre). A kérdésben rejlő egyedi problémákhoz talán a keresés teljesen más modellje révén jutunk el, mely éppen azért ismeretlen, mivel az adott probléma eddig csak ritkán vetődött fel, s megoldása nem járt modellalkotó általánosításokkal.

NYISS!

Tágítsuk ki a keresési tartományt. Lehet, hogy a kérdéshez választott vagy feltételezett szakterület, tudományág túl szűk, a benne felhalmozódó ismeretek nem elegendők a válasz megfogalmazásához.

ALKUDOZZ! (a felhasználó befolyásolása)

Vizsgáljuk meg az adott kérdés vonatkoztatási rendszerét, azt a keretet, amelyben a FELHASZNÁLÓ a problémáját megközelítette. Fogalmazzuk meg azt, hogy szükség esetén tisztázhatjuk vele, mennyiben változtatható az ő kiindulópontja anélkül, hogy az alkudozás eredményeként túlzott torzulások keletkeznének. Az igen-nem válaszok eredményeként az n-edik lépésben megszülethet a kérdésnek a felhasználó számára még elfogadható, a kereső számára pedig megoldható a kérdésnek a megfogalmazása.

TÁROLJ!

Bonyolult kereséskor számtalan forrást nézünk át, míg meg nem találjuk azt, amely a leginkább megfelel a kérdésnek. A közbenső információk is elárulhatnak valamit a leginkább megfelelő információ természetéről, esetleg azt is jelezhetik, hogy az eredeti kérdést zagyván fogalmaztuk meg, s ezért újra kell fogalmazni. Azaz nemcsak arra van szükség, hogy számon tartsunk minden nyomravezető elemet, melyről feltételezhető, hogy

megváltoztathatja a kérdés természetéről vagy a legmegfelelőbbnek tartott információról alkotott egyik vagy másik elképzelésünket. Az is fontos, hogy felhasználjuk a korszerű keresőrendszerek ama képességeit, hogy – automatikusan vagy külön utasításra – tárolják a kérdéseket és megjelenítsék a keresési történetet.

HÖKKENTS!

Előfordulhat, hogy hiába alkalmaztuk a fent felsorolt taktikákat. Ilyenkor talán a leleményesség és a szellemesség segíthet: Találjunk ki valamilyen lehetetlen, eszeveszett vagy legalábbis drámaian új megoldást a problémára! A hökkentés “oldalazó” gondolkodás. Segítségével mintegy rendeződnek a meglévő információk, felismerhetők a hagyományoktól eltérő modellek és teljesen átfogalmazható a kérdés. A bevált sémák meghatározzák a gondolkodás folyamatát, a gondolati “félrelépéssel” vagy “oldalazással” mintegy rápillantunk arra a modellre, mellyel azonosultunk, s ezért megköti a képzeletünket.

VÁLTS!

Változtassunk meg valamit – bármit – a keresési viselkedésben; próbálkozzunk más forrásokkal, más kifejezésekkel, más témakörrel stb. A váltással egyben a megszokott keresési modell is automatikusan törlődik. Az új viselkedés termékeny gondolatokat sugall.

SZŰKÍTS!

Vizsgáljuk meg a kérdést közelebbről oly módon, hogy (1) a kérdés egészéről a kérdés egyik részletére irányítjuk a figyelmünket, vagy oly módon, hogy (2) a kérdés speciális értelmezésére térünk át (akár mindkettőt alkalmazhatjuk).

BŐVÍTS!

Vizsgáljuk meg a kérdést távolabbról oly módon, hogy (1) a kérdés egyik részletéről az egészre irányítjuk a figyelmünket, vagy, hogy (2) a kérdés speciálisabb értelmezéséről egy általánosabb értelmezésre térünk át (akár mindkettőt alkalmazhatjuk).

UGORJ!

Változtassunk a kérdés megközelítésén oly módon, hogy (1) az összetett, több részből álló kérdés egyik részlete helyett egy másik részlet keresésébe fogunk, vagy oly módon, hogy (2) a kérdést más szemszögből vizsgáljuk meg, amely se nem tágabb, se nem szűkebb, mint a korábbi megközelítés, hanem egyszerűen más.

ÁLLJ!

Függesztjük fel a keresést és foglalkozunk valami mással. Persze idővel újra elő kell venni, hátha addigra újabb ötleteink támadnak. Ha viszont reménytelen, ne függesztjük fel, hanem fejezzük be.

3.8 A keresést előkészítő és befejező eljárások

A keresés nem csak a keresőszavak kiválasztásával, a keresőkép szerkesztéséből és a keresési stratégiából áll. Ezek előtt az alapvető műveletek előtt célszerű végiggondolni a tervezett keresést tágabb összefüggések között is.

1. (Ön)interjú

– Alkalmas-e egyáltalán a téma online keresésre?

Nem lehetséges, hogy nyomtatott dokumentumokban, kézikönyvekben, telefonos tájékozódás alapján könnyebben vagy kimerítőbben választ kapunk?

– Mennyire vagyunk jártasak az adott témában?

Pontos elképzelésünk van arról, amit keresni akarunk, vagy inkább általánosságban szeretnénk tájékozódni a témáról? Ettől függően célszerű például kiválasztani az adatbázist, amelyben keresni fogunk, és az információkereső nyelvet. Speciális tájékozódásra a teauruszok, tárgyszavas rendszerek, átfogóbb tájékozódásra például az ETO alkalmasabb. Szükség esetén vonjunk be szakértőt, elsősorban azt, aki a keresést kérte.

– Mit ismerünk a témából? Van-e tudomásunk a témába vágó szakcikkekről, folyóiratokról stb.

A már ismert publikációk fontos fogódzók lehetnek a keresőkérdés első megfogalmazásában, sőt más források megtalálásában is (például hivatkozások révén).

- Milyen szakkifejezéseket ismerünk a témával kapcsolatban?
Célszerű már előre számba venni a szinonimákat, rövidítéseket, idegen nyelvű ekvivalens kifejezéseket, amelyek beépíthetők a keresőképbe.
- Vannak-e a témát korlátozó szempontok?
Lehet, hogy csak meghatározott dokumentumtípus, nyelv, földrajzi terület vagy időszak érdekel. Az ilyen korlátozó szempontok megjelölésének a könyvtári rendszerekben viszonylag nagy választéka van, érdemes róluk már előzetesen tájékozódni, hátha olyan korlátozó szempontok is az eszünkbe jutnak, melyekről nem tudtunk.
- Van-e értelme (vagy van-e lehetőség rá), hogy kereskedelmi szolgáltató adatbázisában keressünk?
Ennek nem csak anyagi, hanem hozzáférési feltételei is vannak. Rendelkezni kell a belépéshez jelszóval.
- Mennyire sürgős a keresés?
Gyakorlatlan kereső sürgős esetben olykor hamarabb célhoz ér, ha segítséget (például szaktájékoztatót vagy információs brókert) vesz igénybe. Ez például akkor fordulhat elő, ha kereskedelmi szolgáltató által kínált adatbázisban nagyon összetett keresést kellene parancsnyelven elvégezni.
- Milyen formában van szükségünk az eredményre?
Elég, ha a képernyőn jelennek meg a találatok, vagy szükségünk van nyomtatott, letöltött, netán faxon elküldött találati jegyzékre? Más és más lehetőségek állnak rendelkezésre bibliográfiai jegyzékek és teljes szövegű dokumentumok esetén.

2. Keresési cél

- Mire akarjuk fölhasználni a keresés eredményét?
- Mennyi találatra számítunk?
- Mi a fontosabb: a nagy teljesség, a nagy pontosság vagy a teljesség és pontosság arányos megvalósítása? (Lásd a 8. fejezetet!)

3. Szolgáltató kiválasztása

- Könyvtári adatbázisban vagy kereskedelmi szolgáltató adatbázisaiban célszerűbb keresni?

A találatok birtokában utóműveletekre is célszerű sort keríteni.

1. Értékeljük az eredményeket
2. Szükség esetén ismételjük meg a teljes keresést más körülmények között.

A fentiekben vázolt megfontolásoknak különösen akkor van jelentősége, ha a feladat mások keresési megbízásainak támogatása vagy teljesítése a könyvtári referenzszolgálat keretében vagy professzionális információközvetítőként (információbrókerként).

Az utóbbi tevékenységre vonatkozóan az Információszolgáltatók Európai Szövetsége (The European Association of Information Services, EUSIDIC), karöltve más szervezetekkel erkölcsi kódexet¹¹ fogalmazott meg, melynek legfontosabb követelményeit az alábbiakban foglaljuk össze:

- Úgy dolgozzék, hogy a munkája alapján a felhasználó kedvező képet alkosson az online keresésről általában.
- Ne tegyen különbséget a felhasználók között.
- Ne ígérjen többet, mint amit teljesíteni tud.
- Tárgyilagosan értékelje a saját munkáját.
- Őrizze meg a tudomására jutott (üzleti) titkokat.

¹¹ Code of Practice for Information Brokers, 1994 <<http://v.hbi-stuttgart.de/~capurro/ethikreader.html#Eusidic>> A kidolgozásában részt vett még az Európai Információkeresők Hálózata (The European Information Researchers Network, EIRENE)

- Ne éljen vissza a tudomására jutott információval.
- Tartsa tiszteletben a szerzői jogokat.
- Legyen elfogulatlan az információforrás kiválasztásában.
- Tartsa tiszteletben a felhasználó kéréseit.
- Tisztességesen tájékoztassa a felhasználót, hogy milyen kritériumok alapján keresett.
- Ne hamisítsa meg a keresés eredményét.
- Még a munka előtt tájékoztassa a felhasználót a lehetséges költségekről.
- Folyamatosan képezze magát.

3.9 Az információkeresés lélektani összefüggései

3.9.1 Az emberi tényező szerepe

A tudományos kommunikáció rendszerén belül világosan megkülönböztethetők az **információk létrehozói**, az **információk**, az **információátvitel csatornái** (beleértve az emberi feldolgozókat és közvetítőket is) és a **címzettek** (az információk felhasználói). Annak, hogy az információk az információátvitel csatornáin keresztül gyakran töredékesen és zajok által zavarva jutnak el a címzettekhez, illetve, hogy a címzettek sok esetben elégtelenül és torzítva használják fel a hozzájuk érkezett információkat, több oka lehet, melyek közül a fontosabbak:

1. az információforrások ismeretének hiánya vagy az elégtelen szakismeret – a felhasználó nem tudja vagy nincs tudatában annak, vagy nem hiszi, hogy
2. a számára szükséges információ valahol már megszületett és rögzítették; a kellő műveltség hiánya – a felhasználó nem ugyanazon a műveltségi szinten van, mint amely szinten az információ keletkezett (az információ létrehozójának, közvetítőjének vagy a felhasználójának eltérő a műveltsége); ezt egyes esetekben tekinthetjük nyelven belüli akadállyal is;
3. a nyelvek közötti akadály – a közvetítő vagy a felhasználó nem ismeri (nem ismeri jól) az információ megfogalmazásához, rögzítéséhez használt idegen nyelvet, illetve szaknyelvet;
4. a technikai akadály – technikai okokból nem hozzáférhetők az információk.

A négy akadály közül az első három lelki, intellektuális és nyelvi (a továbbiakban – mivel szorosan összefüggenek a lelki/tudati és a nyelvi szinttel, az egyszerűség kedvéért összefoglalóan pszicholingvisztikai természetűnek fogjuk fel őket), csak a negyedik műszaki. Ebből következik, hogy hiába vannak megfelelő műszaki eszközök, ha a kommunikációnak pszicholingvisztikai akadályai vannak. A tudományos kommunikáció, s vele az információkeresés folyamatát eme akadályok szempontjából is célszerű elemezni.

Noha mára az információátvitel csatornáit elsősorban a számítástechnikai eszközök alkotják, nyilvánvalónak kell tehát lennie, hogy **e csatornák meghatározó eleme továbbra is az ember, akinek információközvetítő munkáját – az osztályozást és az indexelést meg a keresést – éppen úgy befolyásolják a pszicholingvisztikai tényezők, mint az információk alkotóiét és felhasználóiét.** Vagyis az információközvetítés éppen úgy alkotómunka, akár a másik kettő.

Az információkeresésről szóló 2.1 fejezetben rámutattunk arra, hogy az indexelés és osztályozás – az információátvitel leglényegesebb intellektuális mozzanata – nem más, mint valamely természetes nyelven írt dokumentum tartalmának lefordítása valamely szabványosított információkereső nyelvre.

A fordítást adott “rendező rendszer” (információkereső nyelvi szótár: mint például osztályozási rendszer, tárgyszójegyzék, teaurusz) segítségével végzik. Ezek alapján készül el a dokumentumkép a dokumentum tartalmáról. E dokumentumkép segítségével válhat a keresés algoritmizált eljárássá.

Függetlenül attól, hogy az indexelés, illetve osztályozás milyen “rendező rendszer” igénybevételével folyik, valamint feltételezve, hogy az indexelők, illetve osztályozók a lehető legpontosabban betartják az általuk alkalmazott “rendező rendszer” szabályait, két indexelő ugyanarról a dokumentumtartalomról eltérő dokumentumképet alkothat, aminek okai összefüggnek a fentebb tárgyalt három pszicholingvisztikai problémával.

A dokumentum, az indexelő/osztályozó és a keresőkép közötti háromszögben a következő, egyelőre végleges válasz nélkül maradó kérdések vetődhetnek föl¹²

- formalizálható-e a dokumentum tartalmi értelmezésének a folyamata?
- formalizálható-e a természetes nyelv bármiféle keresőnyelvre történő fordítása?
- lehetséges-e az olvasók valamennyi kérdését előre látni, illetve pontosan leírni?
- az indexelés/osztályozás melyik módja formalizálható, és melyik nem?
- létezik-e elvileg olyan indexelési/osztályozási mód, amely algoritmizálható?

A fenti kérdésekre (az utolsó kivételével, melyre ma az automatikus indexelési és osztályozási eljárások “válaszolnak” a maguk korlátozott módján) még alig adhatunk tudományosan megalapozott választ. Annyi valószínű, hogy a számítógép nem válthatja fel az embert, viszont megnövelheti a különféle szellemi tevékenységek hatékonyságát. Ezzel kapcsolatban érdemes emlékeztetni Jesse Shera talán túlzott, de a valóság lényeges elemét megragadó következő intelmére: “A könyvtári-információs gyakorlatban még egyetlen jelenség sem váltott ki olyan széles közérdeklődést, mint a gépesített információkeresés, de egy sem volt improduktívabb nála, minthogy a figyelem homlokterébe a gépek kerültek, nem pedig az emberi, a logikai, a nyelvészeti stb. aspektusok.”

3.9.2 Az információkeresés kognitív modellje

A könyvtáros információkereső tevékenységében fontos szerepet játszik **műveltsége, keresési tapasztalata és munkaterületének adottságai**. A szakmai műveltség hatékonyságát befolyásolja a könyvtáros általános műveltsége is, és a keresési tapasztalatok hozzájárulnak meglévő műveltségének fejlődéséhez.

Ez a műveltség ismereteket jelent, és az ismereteknek fogalmi természete, struktúrája, rendszere van, mely – különösen kommunikációs szinten – elválaszthatatlan egyrészt a nyelvi műveltségtől, másrészt a lélektani állapottól, konstitúciótól. Azt a szakterületet, amelynek tárgya a gondolkodás, a fogalmak és a lelki működés összefüggése, kognitív lélektannak nevezik. Az információkeresés kognitív lélektani elemzésére történtek már kísérletek, a dán Peter Ingwersen nevéhez egy kognitív modell kidolgozása fűződik¹³.

A modell alapja maga is egy (pontosabban három) modell. Az előző fejezetben tárgyalt információ-létrehozók, -közvetítők és -felhasználók mindegyike valamilyen elképzeléssel rendelkezik a világról. Ez az elképzelés a műveltsége, ismeretei (a tudása) révén alakul ki, mely az egyéni és kollektív (társadalmi) tapasztalat, a nevelés és az oktatás során alakul ki; fogalmak, a belőlük felépülő ismeretek rendszeréből áll, és **a világ individuális kognitív modelljének fogható föl**.

Az ismeretstruktúrák (a kognitív világmodellek) afféle “térképek” együtteséhez hasonlíthatók, melyek például az iskolák tantermeinek falain függenek. Egy-egy tárgykörre vonatkozóan egyszerre több “térkép” is érvényes, még egyetlen személy tudatában is, s ezek a “térképek” a tudatban attól függően (is) rendeződnek, hogy melyik az éppen uralkodó szempont. Az információkereső rendszer résztvevői közötti kölcsönhatások, interakciók során más és más

¹² Szöveggyűjtemény, 2. kötet, “Heurisztikus és lélektani megközelítés” című rész, p. 359–365.; a Shera-idézet helye uo., p. 360.

¹³ Szöveggyűjtemény, 2. kötet, “Kognitív szempontú elemzés”, p. 361–364.

“térképek” kibontása válhat szükségessé. Az is előfordulhat, hogy ezek a térképek az interakciók során módosulnak.

E kognitív modellek felől nézve az eredményes információkeresés feltétele az, hogy az információk létrehozóinak (szerzők, közreadók), közvetítőinek (kezdve a számítástechnikai szakembertől, az indexelőkn/osztályozókon, az információs brókereken át a referenszszolgálat szakembereiig) és felhasználóinak kognitív világmodelljei részben vagy egészben, kisebb vagy nagyobb mértékben összehangolódnak.

Tökéletes összehangolódásra nem lehet számítani, és nem is lenne szükséges, mert akkor információigények se merülhetnének föl. A kognitív világmodellekben azoknak a részeknek az összehangolódása szükséges, melyek közvetlen szerepet játszanak az információk továbbításának, feldolgozásának és keresésének folyamatában. A többi, lényegesen nagyobb világmodellrészekben jelentős eltérések is lehetnek, noha a túl nagy eltérések esetében előfordulhat, hogy az eredményes információközvetítés válik – átmenetileg – lehetetlenné.

Az egyének eme föltételezett kognitív világmodelljei szorosan összefüggenek az őket körülvevő kultúra (korábban elsősorban vallási, ma elsősorban tudományos) paradigmáitól. Ezek a paradigmák az adott korszakban elfogadott általános (vallási, tudományos) tételek együttese. Ma a kor tudományos világképét alkotják és meghatározó a hatásuk a tárgyalt egyéni kognitív világmodellekre.

Az Ingwersen-féle modellben a hangsúly az információkeresési folyamaton van. A modellben az említett három résztvevő – az alkotók, a közvetítők és a felhasználók – világképével és ezek célszerű mértékű összehangolásával kell számolni ahhoz, hogy információkereső helyzetben sikeresek legyenek. A három résztvevő közül a közvetítőkre hárul a legfontosabb szerep: az indexelők és osztályozók kognitív világmodelljei meghatározó szerepet játszanak a dokumentumképek és a keresőképek, s rajtuk keresztül a keresési stratégia kialakításában. Következésképp az ő világmodelljeiknek kell a legnagyobb mértékben alkalmazkodnia az információkeresési folyamatok során annak érdekében, hogy a folyamat eredményes legyen.

A legkézzelfoghatóbb formában ez az alkalmazkodás abban nyilvánul meg, hogy az információk és a keresőkérdések feldolgozásához használt információkereső nyelveket és osztályozási rendszereket (a tárgyszójegyzékeket, teauruszokat, “szabad tárgyszavakon” alapuló online mutatókat és osztályozási táblázatokat) a lehető legnagyobb mértékben közelíteni kell azokhoz a kognitív világmodellekhez, melyekkel – az alkotók kognitív világmodelljeinek “ismeretében” – a felhasználók rendelkeznek. (Az idézőjel azért szerepel, mert a szóban forgó ismeretek jelentős részben intuitív folyamatok eredményeként alakulnak ki.) Más szóval és nagyon gyakorlatias megközelítésben például **egy teaurusz szókincse lehetőleg igazodjék a rendszert használó legtájékozatlanabb felhasználó igényeihez, miközben lehetővé kell tennie a legműveltebb alkotó műveinek tartalmi feltárását is.**

A 18. ábra jobb oldalán a felhasználói igény megfogalmazásának folyamata látható. Ha a felhasználó kognitív világmodellje és az őt körülvevő kultúra paradigmái között az eltérések miatt a felhasználó hiányosnak ítéli az ismereteit, bizonytalansággal járó lelki feszültség keletkezik benne (mondják rendellenes tudatállapotnak is). Erre az állapotra jellemző az információs igény. E tudatállapottól a megfogalmazott keresőkérdésig bonyolult tudati folyamatok zajlanak le, s közöttük legalább egy fordítási folyamat is, melynek során a felhasználó a gondolatait, s vele kérdéseit mondatok formájában megfogalmazza.

Az ábra közepén a közvetítő könyvtáros problémamegoldó helyzetben van. A kérdés megértéséhez egyrészt aktivizálnia kell a kognitív világmodelljéből következő szakismereteit, másrészt meg kell értenie a felhasználó kérdését. Ez utóbbihoz kisebb-nagyobb mértékben képet kell alkotnia a felhasználó kognitív világmodelljének a kérdéssel összefüggő részeiről is (már amennyire ez lehetséges), amelybe beletartozik annak felderítése is, hogy milyen helyzetben igényel választ a felhasználó. Ennek érdekében a legcélszerűbb, ha interakció (“párbeszéd”) alakul ki közöttük és a kérdező között; ezáltal a könyvtáros kiinduló képe a felhasználóról és kérdéséről finomodik, ami érinti a könyvtáros kognitív világmodelljét, s adott esetben meg is változtathatja, még ha rendkívül kis mértékben és elsősorban szakmai vonatkozásban. (Megint csak nagyon gyakorlati

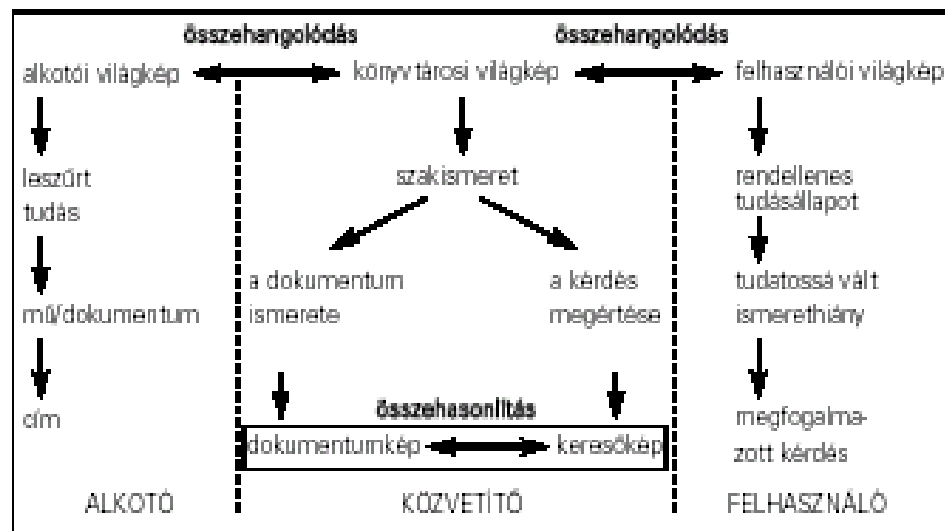
megközelítésben például abban, hogy rájön: kezdő [olykor nem csak kezdő] kutatók számára célszerű a tárgykörben létező referenz-dokumentumokat – bibliográfiákat, kézikönyveket stb. – is ajánlania, amikor ezek nyomban folyóiratcikkek keresésébe akarnak kezdeni.)

Mindezek alapján jöhet létre olyan keresőkép, és alakulhat ki olyan keresési stratégia, melynek segítségével optimálisan végezhető el a tárolt dokumentumképekkel az összehasonlítás.

A közvetítő könyvtáros lényegében hasonló helyzetben van az ábra bal szélén szereplő alkotókkal is, azzal a nem elhanyagolható különbséggel, hogy velük a személyes kapcsolatfelvételre, s ezzel helyzetük és kognitív világmodelljük közvetlenebb megismerésére nincs sok lehetőségük. Ezeket az ismereteket szinte kizárólag az alkotók feldolgozandó műveiből és annak a szakterületnek a többi műveiből szerezhetik meg, melyhez a feldolgozandó mű is tartozik. Mindezek alapján jöhet létre olyan dokumentumkép, amely összhangban van az alkotók művet létrehozó intencióival és vele az alkotók kognitív világképével.

A bal szélén szereplő alkotó és az általa célba vett, elképzelt felhasználókra vonatkozó ismeretei tömören a dokumentum címében érvényesülnek: olyan címet választ, melyről feltételezi, hogy – amellet, hogy kifejezi az ő kognitív világmodelljének részét képező mondanivalóját – megfelel a felhasználó kognitív világmodelljének. (Ezt persze az információkeresés tárgyalt kognitív modelljének nyelvén fogalmazzuk meg ilyen körmondfonatan: a gyakorlatban “gondolta a fene” – hogy Arany Jánost idézzük. Az alkotó praktikusabban talán inkább úgy fogalmazhat, hogy igyekszik érthető, ugyanakkor kellő mértékben figyelemfelkeltő, egyúttal tömör címet adni a művének.) Azt is mondhatjuk, hogy a mű címe az alkotó által megfogalmazott “dokumentumkép”.

*



18. ábra. A könyvtári kommunikációs rendszer kognitív modellje Ingwersen szerint

*

A 18. ábrán látható modellből kiolvashatók azok a lehetséges következmények és bonyodalmak is, amelyekkel számolni kell, ha az emberi közvetítőt “intelligens” információkereső rendszerrel – például automatikus információkereső rendszerrel – akarják helyettesíteni. Számolni kell azzal, hogy a modellben reprezentált, a valóságban rendkívül bonyolult lépések és interakciók algoritmizálása csak nagyon korlátozottan hajtható végre.

Az információfeldolgozás és -keresés elméleti és gyakorlati szakemberei előtt új feladatok jelentkeznek, ha ismeretalapú (tudásbázisú) információkereső rendszereket akarnak létrehozni: vizsgálniuk kell a tudati információfeldolgozást is, hogy némiképp “intelligens” online keresést támogató rendszereket tervezhessenek.

3.9.3 Keresési helyzetek

A közvetítő és a felhasználó közötti keresési helyzeteket általában az ötleteken, tapasztalatokon alapuló, rávezető heurisztikus eszközöket használó **nyitott keresés és a szimmetrikus**

megbeszélési helyzet jellemzi. Ez utóbbi azt jelenti, hogy – együtt a dokumentumokkal és a segédletekkel – a közvetítő könyvtáros és a felhasználó között oda-vissza lejátszódó párbeszédre (interakciókra) kerülhet sor, azaz a felhasználó az információforrás szerepét játssza, és javítja a közvetítő képét a keresési igényről. A közvetítőt az motiválja ebben, hogy igyekszik az igény megértéséhez sok releváns információhoz jutni, mielőtt a keresési probléma tényleges megoldásába egyáltalán belekezdene.

Ugyanakkor feltételezhető **a félkötött keresés és az asszimmetrikus megbeszélési helyzet** is. Ebben a felhasználónak passzívabb a szerepe, a közvetítő könyvtáros inkább a felhasználó által konkrétan megadott kérdésekre és a belőlük következő fogalmakra támaszkodik, nem vagy alig kérdez vissza, mert nem akar vagy nem tud különösebben módosítani rajtuk, és így kísérli meg közvetlenül megtalálni a releváns dokumentumot.

A teljesen kötött, algoritmikus keresés és asszimmetrikus megbeszélési helyzet, melyben semmiféle interakció nem játszódik le, elsősorban az automatikus indexelésre és osztályozásra jellemző (bár emberi mulasztás vagy tudatlanság miatt a “közvetítés” ennél a keresésnél is előfordulhat). Átlagos munkaköri körülmények között, amikor nem kell túlhajszoltsággal vagy negatív lélektani beállítódással számolni, Ingwersen kísérleti eredményei szerint ritkán fordul elő, hogy feltett kérdésre a közvetítő, ha nem érti a kérdését, mégis automatikusan keresni kezd. Ez talán annak is tulajdonítható, hogy a felhasználóknak nem lebecsülhető a hatása a közvetítő könyvtárosokra. Amikor a keresés azonnal, kötött és asszimmetrikus kiindulás alapján nem vezet eredményre, a könyvtárosok utat engednek a visszacsatolásnak, és interakcióba lépnek a felhasználóval.

A felhasználó és az online katalógus közötti közvetlen interakció viszonylag a legalacsonyabb színvonalú, a felhasználók túl egyszerű stratégiákat használnak, amin az online katalógusok sugói se sokat segíthetnek, részben hiányosságaik miatt, részben azért, mert a felhasználó nem használja őket (mintha 25–30 éves kor fölött sokan már nem lennének hajlandók újat tanulni). Peter J. Vigil¹⁴ ezzel összefüggésben mutatott rá arra, hogy a felhasználók a VAGY és az ÉS műveleteken kívül más logikai műveletet jóformán nem használnak. A logikai műveletek közül ugyanis a NEM művelet fogható föl a legnehezebben. Ez pszichológiai tény, a tagadó mondatok több gondolkodási időt igényelnek, mint az állító mondatok. Ezért a negációt a felhasználók a redundáns hivatkozások kiküszöbölésére sem használják, holott erre a célra elengedhetetlen volna.

Az online információközvetítők számára a nyitott keresési módok alkalmazása a leghatékonyabb, mert ebben összekapcsolódnak a keresés elején alkalmazott heurisztikus módszerek a későbbi formálisabb megoldásokkal, melyekben akár az információkereső rendszerben meglévő, beépített algoritmusok is kihasználhatók a keresési stratégia folyamatában, hogy megvalósuljon a keresőkép és a dokumentumkép optimális összehasonlítása.

A lélektani elemzések arra utalnak, hogy az információközvetítés döntő mozzanata az esetleg rosszul megfogalmazott kérdés mögött meghúzódó információigény pontos megértése, hogy ezt követően a kérdést átalakíthassák az információkereső rendszer struktúráinak megfelelő módon és nyelvre. Ehhez a használónak segítségre és ismeretekre van szüksége.

Gyakori tapasztalat, hogy az átlagos felhasználó az online olvasói katalógusok használatában is nagyon tájékozatlan, ugyanakkor ebből következően önmagát megrövidítően magabiztos. Például anélkül, hogy ellenőrizné, magyar címszavak szerint keres olyan állományban, melynek dokumentumtétellei idegen nyelven íródtak, és az eredmény láttán rossznak tekinti a rendszert. Az ilyen “önálló” információkeresést végző olvasó tapintatos “kézenfogása”, kiszolgálása és egyben tanítása is (vagyis ami a referenz-könyvtáros dolga) út és segítség, még ha nem is magának a keresésnek a közvetítéséről, hanem csak a tanításáról van szó. Nem állítható tehát, hogy az alternatíva vagy a közvetítő, vagy a közvetítő nélküli felhasználó.

¹⁴ Peter J. Vigil: Az online keresés pszichológiája. In: Tudományos és Műszaki Tájékoztatás, 1984, 31. évf., 11. sz., p. 324–327.

4. Adatbázisok, szolgáltatás

4.1 Távoli és helyi hozzáférés¹⁵

Ameddig a személyi számítógépek és velük a helyi hálózatok nem terjedtek el, az adatbázisokhoz kizárólag az adatbázist kezelő nagyszámítógépeken, illetve a hozzájuk kapcsolódó távközlési rendszereken keresztül lehetett hozzáférni. Nagyszámítógépből azonban intézményenként többnyire egyetlen egy állt rendelkezésre, helyi használatra csak ennek a monitorjait lehetett igénybe venni, ami alkalmatlan volt az információkeresés ad hoc és tömeges igényeinek kielégítésére. A gyakorlatban csak az offline és a távoli online keresőszolgáltatások rendszeres használata volt a jellemző, a nagyszámítógépeken keresztül. Ahány adatbázis, annyiféle információkereső nyelvre, parancsnyelvre és egyéb jellemzőre kellett számítani, amiből következett, hogy elsősorban a professzionális információközvetítők tudtak élni a lehetőségekkel. Az online hozzáféréseken és a távoli használaton ebben az időszakban jóformán ugyanazt értették.

A nyolcvanas évek közepétől ugrásszerű változás állt be: a helyi hálózatba kapcsolt személyi számítógépek és terminálok használója bármikor hozzáférhetett a központi gépen kezelt helyi – tehát “saját” – adatbázishoz. Ennek köszönhetően jelentek meg – általában a szintén ekkor elterjedő integrált, számítógépes könyvtári rendszerek “termékeként” és részeként – többek között az online [nyilvános hozzáférésű] katalógusok (angol rövidítéssel: OPAC; Online Public Access Catalogue). Mindez azzal járt, hogy jelentősen megnőtt azok száma, akik az online szolgáltatásokat igénybe vehették, és jól eligazodtak saját helyi információkereső rendszerükben. A távoli online hozzáférés ezekről a személyi számítógépekről is a helyi hozzáféréshez képest körülményesebb volt, és speciális szakértelmet igényelt.

A kilencvenes évek elején újabb jelentős változás körvonalai rajzolódtak ki. Az internet megjelenésével egyrészt a felhasználóbarát kezelőfelületek révén a távoli online kereső szolgáltatásokhoz való hozzáférés is könnyebbé vált, másrészt a helyi adatbázisok most már nemcsak a helyi, hanem gyakran a távoli hozzáférés számára is rendelkezésre álltak. Világméretben elterjedt az internet használata. A lehetőségek eme rendkívüli “demokratizálódásával” párhuzamosan technikailag egyszerűsödött és nem utolsósorban olcsóbb lett a távoli hozzáférés. Az egyes adatbázisok specifikus, nem mindenki által rögtön megismerhető, elágazó, lassú (ennél fogva drága) paranccsal vagy menüvel vezérelt keresőrendszerét az interneten alkalmazott web gyorsabb, ún. ablaktechnikát alkalmazó grafikus felülete váltotta föl. Ez az ablaktechnika a későbbiekben megjelent a helyi rendszerek keresőfelületein is.

Mára gyakorlatilag értelmét veszítette a távoli és a helyi hozzáférés megkülönböztetése, mivel az adatbázisok számára általában kialakítják azt a felületet, melyen a web keresőeszközeivel a hálózaton keresztül keresni lehet. Gyakorlatilag minden online távoli hozzáférés az interneten keresztül valósul meg. Információk feldolgozását továbbra is végzik a helyi, különböző szoftvereket (mint az ALEPH, az OLIB, a CORVINA, AMICUS stb.) alkalmazó rendszerekben és így a könyvtári adatbázisokban, az online olvasói katalógusokban továbbra is keresni lehet az adott könyvtáron belül. A keresések számottevő részét azonban már “távolról”, az adott könyvtáron kívülről kezdeményezik (más könyvtárból, munkahelyről, otthonról, vagy az ún. webkávézóknak nevezett nyilvános helyiségekből). Számos könyvtári adatbázis már eleve úgy készült, hogy csak a weben keresztül lehet keresés céljából hozzáférni.

Mindez vonatkozik a kereskedelmi jellegű, távoli hozzáférésű adatbázisokra is. A professzionális kereső ugyan továbbra is az adott adatbázis speciális parancsnyelvét kénytelen használni, ha azt akarja, hogy keresése eredményes, és főleg kifizetődő legyen, de a keresést már közvetlenül saját személyi számítógépén valósíthatja meg, a hozzáférésre pedig az internet eszközeit használja.

4.2 Adatbázisok típusai

¹⁵ Szöveggyűjtemény, 2. köt. p. 382–425.

Az adatbázisokat a szolgáltatott információk szempontjából a Carlos A. Cuadra által kialakított taxonómia alapján szokták csoportosítani.

REFERENZ ADATBÁZISOK – az eredeti, elsődleges forrásokra vonatkozó ún. másodlagos információkat tartalmaznak. Két fajtájukat különböztetik meg:

BIBLIOGRÁFIAI ADATBÁZISOK – az elsődleges források a kiadott vagy nem kiadott dokumentumok (könyvek, időszaki kiadványok, folyóiratcikkek, szabadalmak, térképek, zeneművek stb.). A másodlagos információk a rájuk vonatkozó dokumentumtétel (dokumentumleírások, bibliográfiai tételek). E tételekhez kapcsolhatnak referátumokat, annotációkat, tömörítvényeket stb.

Ilyen adatbázisok a könyvtárak online olvasói katalógusai, a bibliográfiák, cikkadatbázisok, továbbá a különféle szakterületi dokumentumrekordokat tartalmazó adatbázisok (például a MEDLINE, vagy nálunk a PRESSDOC adatbázisa).

FORRÁSTÁJÉKOZTATÓ ADATBÁZISOK – az elsődleges információforrások (személyek, szervezetek, szolgáltatások, folyó kutatások (“projektek”) stb. (tehát nem dokumentumok) adatait tartalmazó és kereshetővé tevő adatbázisok. A másodlagos információk a rájuk vonatkozó rekordok (leírások), melyek alapján elérhetők.

Ilyen adatbázisok például a cím- és céginformációs adatbázisok (név- és címtárak, cégkatalógusok), telefonkönyvek, termékinformációs adatbázisok (melyekben az előállított áruk, szolgáltatások elérésére vonatkozó adatok, a szolgáltatók, cégek, egyéb szervezetek adatai) szerepelnek (például a KSH cégnyilvántartása, az “Arany Oldalak” szakmai telefonkönyv weben, wapon [mobiltelefonon] és CD-ROM-on elérhető adatbázisa).

FORRÁS ADATBÁZISOK – elsődleges adat- vagy információforrások, amelyek az eredeti forrás tartalmát (számszerű vagy szöveges adatait, teljes szövegét) szolgáltatják. Három fajtájukat különböztetik meg:

NUMERIKUS ADATBÁZISOK – eredeti statisztikai vagy más numerikus adatokat (például termelési, ipari, fogyasztási, költségvetési, pénzügyi, választási, demográfia adatokat tartalmaznak a legkülönbözőbb (például idősoros, területi) rendezettségben.

Ezeket az adatbázisokat sokszor adatbankoknak nevezik; egy részük nem nyilvános. A legismertebbek a statisztikai hivatalok adatbázisai, a demográfiai adatbankok, a lakossági címnyilvántartás, az igazgatási (minisztériális és önkormányzati) adatbázisok (például a Defense Data Bank, mely az USA védelmi minisztériumának gazdálkodási adatait tartalmazza) és az ipari termelés és szolgáltatás adatait tartalmazó adatbankok (például az USA mezőgazdasági adatait heti, havi és negyedéves idősorokban szolgáltató National Agricultural Statistical Service, a termelés, fogyasztási és népességi adatokat tartalmazó PTS Time Series).

SZÖVEGES-NUMERIKUS ADATBÁZISOK – eredeti szöveges és számszerű adatokat felváltva tartalmaznak.

Életrajzi, vállalati, politikai stb. adatbázisok, olykor kézikönyvként nyomtatott formában is léteznek. Ilyen adatbázisok például a Pályázatfigyelő, a CompAlmanach Kft. Ki mit gyárt? vállalati katalógusa, a vegyi anyagokról szóló információkat szolgáltató CHEMSEARCH.

TELJES SZÖVEGŰ ADATBÁZISOK – eredeti szöveges dokumentumokat tartalmaznak.

Ilyen adatbázisok az elektronikus könyvtárak (például az OSZK-ban fenntartott Magyar Elektronikus Könyvtár), a digitalizált enciklopédiák (például a Pallas Nagy Lexikona CD-ROM-on), elektronikus folyóiratok és lapok cikkeinek adatbázisai (például a Magyar Hírlap vagy a Times archívumai), a bírósági határozatok, jogi döntések gyűjteményei stb.

4.3 Adatbázis-szolgáltatás és könyvtári tudásmenedzsment

Az online hozzáférhető adatbázisok szolgáltatói vagy terjesztői olyan szervezetek, amelyek a felhasználók vagy az információközvetítők részére online hozzáférést biztosítanak a géppel

olvasható formában elérhető adatbázisokhoz. Az adatbázis-szolgáltató beszerzi vagy egyes esetekben maga építi azokat az adatbázisokat, amelyekről úgy véli, hogy használói igénylik. Olyan szoftverrendszereket készít és fejleszt vagy alkalmaz, amelyek lehetővé teszik a keresést ezekben az adatbázisokban. Nagyszámítógépeket lízingel vagy vásárol, vagy azokat – a kiegészítő berendezésekkel együtt – karbantartja. Dokumentációt készít, amely bemutatja, hogyan lehet a rendszer segítségével a leghatékonyabban és gazdaságosabban használni az adatbázisokat.

Az adatbázis-szolgáltatók egy része kereskedelmi szervezet. Ezek díjat számítanak fel a nyújtott szolgáltatásért. Rendszerint a díj nagyobb részét az adott keresésre fordított idő függvényében határozzák meg, amelyet a kapcsolati idővel jellemeznek, vagyis azzal az idővel, amelyben a kereső terminálja közvetlen kapcsolatban állt a gazdaszámítógéppel. Az 4. táblázat példaként néhány kereskedelmi adatbázis-szolgáltató adatait tartalmazza.

*

4. táblázat. Néhány jelentősebb adatbázis-szolgáltató.

*

Az adatbázis-szolgáltatók és az adatbázis-előállítók megkülönböztetéséről szót kell ejtenünk. Az adatbázis-szolgáltatóknál elérhető adatbázisok előállítóinak egy része önmaga is biztosít kereső szolgáltatást is (mint a National Library of Medicine vagy az OCLC Inc.). A szolgáltatók többsége ezzel szemben legfeljebb egy-két adatbázist épít azok közül, amelyekhez hozzáférést nyújt. Gyakoribb, hogy gazdasági szempontoktól vezérelve, a kereskedelmi cégek megvásárolják azokat az adatbázisokat, amelyek, reményeik szerint, hasznot hoznak a számukra. Egyes adatbázisok, mint az ERIC, a Medline, az Academic American Encyclopedia és a BIOSIS, több kereső szolgáltatáson át is elérhetők. Más adatbázisoknak egyetlen szolgáltatóval van kapcsolatuk. Ismét mások, például az Institute for Scientific Information vagy a National Library of Medicine által készített adatbázisok, az adatbázis-készítők, és más adatbázis-szolgáltatókon (például a BRS és az SDC) keresztül is elérhetők.

Az igazán eredményes kereséshez nagyon jól kell ismerni mind az adatbázisok, mind a szolgáltatók tulajdonságait. Ezzel összefüggésben szót kell ejtenünk például arról, hogy ugyanazt az adatbázist hálózaton, illetve CD-ROM-on keresztül lekérdezve nem feltétlenül ugyanazt az eredményt kapjuk. Egy szolgáltatón belül is különbség lehet az eredményekben, ha a professzionális keresés számára biztosított hozzáférést vagy valamelyik egyszerűbb webes felülettel vezérelt és adott esetben ingyenes hozzáférést vesszük igénybe.

Tapasztalt információközvetítők a megmondható, mekkora különbségek tapasztalhatók az eredményekben, ha például a Medline adatbázisában a SilverPlatter szolgáltató igénybevétele CD-ROM-on, ugyanezt a szolgáltatót igénybe véve hálózaton, vagy a Medline OPAC-ján, illetve az NLM PubMed webes felületén vagy a DIALOG webes felületén (DialogWeb) vagy a DIALOG hagyományos felületén (DialogClassic) stb. keresztül keresünk. Viszonylag egyszerűen megvalósítható élmény összevetni, milyen találatok kaphatók például ugyanabban a tárgykörben, ugyanazokra a szakfolyóiratokra korlátozva a keresést a PubMedben, és mondjuk a Google általános indexelő szolgáltatásán keresztül.

Tapasztalt információközvetítők azt is tudják, hogy bizonyos dokumentumtípusokat bizonyos szolgáltatókon keresztül eleve nem érdemes keresni. Az esetek többségében viszont ugyanazt a keresést célszerű több elérési úton keresztül is elvégezni, mert például a PubMed szolgáltatja ugyan a legfrissebb tételeket a keresett betegségről, de gyógyszerek szempontjából a Chemical Abstracts adatbázis a legteljesebb. Az is veszteséghez vezethet, ha egy orvos csak a Science Directen keresztül keresne a Medline adatbázisában, de a PubMeden keresztül nem, mert adott esetben csak nagy késéssel találná meg a nem Elsevier-lapokban megjelent cikkeket.

Az adatbázis-szolgáltatók másik – nem kevésbé jelentős – részét a könyvtárak alkotják. A jelentősebb könyvtáraknak – amellet, hogy saját online katalógussal, bibliográfiai adatbázisokkal, CD-ROM-adatbázisokkal rendelkeznek – külső adatbázisokhoz hozzáférési jogosultságuk is van.

Ezért a végfelhasználók szempontjából valójában a legfontosabb adatbázis-szolgáltatók. Versenyképességüket növeli, hogy túlnyomórészt közhasznúak, bárki számára és – egyes szolgáltatások kivételével – ingyen elérhetőek. A könyvtárak olyan szolgáltatók, amelyeknek döntő szerepük van nemcsak a kultúra, a műveltség terjesztésében, hanem a tudományos munka támogatásában is.

A számítógépes információfeldolgozás és -tárolás, az elektronikus dokumentumok és az internet elterjedésével a könyvtárak illetően szerepe nem hogy csökkenni, hanem növekedni fog.

A vállalati szférában alkalmazott tudásmenedzsment a könyvtárak számára is fontos kérdés. Az egyik legrövidebb meghatározás szerint a tudásmenedzsment nem más, mint az intézményi szellemi tőke növelését célzó törekvések összessége. E területen a könyvtári feladatok abban állnak, hogy

- tudatosítsák saját munkatársaik és a közvélemény számára, hogy milyen információkat birtokolnak és képesek szolgáltatni; szervezzék meg a könyvtári munka tanulságainak és tapasztalatainak megosztását a munkatársak között;
- adatbázis-szolgáltatásaikat összefogott, rendezett szervezeti és technikai keretek között kínálják fel;
- csökkentsék használóiknak azokat a nehézségeit, melyet a nagy tömegű, emészthetetlen keretek között áradó információ okoz, és ugyanakkor tegyék lehetővé, hogy az olvasók ismerethiányaikat célzott és támogatott formában feloldhassák.

Mindehhez szükség van ún. tudástérképre. A tudástérkép nem más, mint annak ismerete, hogy adott intézményen belül mikor és hova (kikhez) érdemes fordulni. A legegyszerűbb tudástérkép az intézmény szervezeti felépítése. Ha belegondolunk, hogy különösen egy nagy könyvtárban mennyi és milyen sokféle tudás testesül meg az adatbázisokban és a rájuk támaszkodó könyvtári szolgáltatásokban, máris érzékelhető, mennyire fontos, hogy a könyvtár munkatársai saját, intézményen belüli technikai és szakmai lehetőségeikkel mintegy “térképszerűen” tisztában legyenek.

Tévedés azt hinni, hogy e feladatok megoldásában kizárólag a “megfelelő” információtechnológia segíthet. A technológia csak a könnyebb elérhetőséget biztosíthatja. A tudásmenedzsment feladata éppen az, hogy minden – nem csak számítástechnikai – eszközt felhasználjon az eredményes könyvtári adatbázis-szolgáltatás céljából. A kereskedelmi adatbázis-szolgáltatókhoz képest a könyvtárak lényegesen többet tehetnek a szolgáltatandó információk személyes tudássá átalakításában. E téren a könyvtáraknak katalizáló szerepe van, ami nincs még eléggé tudatosítva.

Az egyik legjelentősebb fejlemény a könyvtári eredmények sikeres “menedzselésében” az ún. digitális referenz. (A magyar “Mit hol?” kérdőívet alkalmazó tájékoztatószolgálat elvileg ebbe az irányba mutat <<http://mit-hol.oszk.hu>>.) Egyik fajtája az elektronikus levelezésen, másik, rohamosan fejlődő fajtája az online “csevegésen” (chat), a “csevegőszobán” (chat-room) alapul. A könyvtárak virtuális referenz-szobát nyitnak, ahol több olvasó is tartózkodhat egyszerre és faggathatja a könyvtárost (természetesen “négyszemköt is”). A fejlettebb változatokban a könyvtáros megfelelő program segítségével online szimulálja azt a helyzetet, mintha ő meg az olvasó a referenz-pultnál állnának. Legismertebb megvalósulása a Kongresszusi Könyvtár égisze alatt létrehozott együttműködő hálózati referenz-szolgálat (Collaborative Digital Reference Service, CDRS, <<http://www.loc.gov/rr/digiref/>>).

A könyvtárak adatbázisaiban és az általuk hozzáférhető külső adatbázisokban fölhalmozott “hallgatag” tudás felszínre hozására, mások számára használhatóvá tételére éppen a könyvtári szervezet az a történetileg kialakult forma, mellyel a pusztán adatokat szolgáltató kereskedelmi szervezetek azért nem versenyezhetnek, mert a könyvtárakkal szemben végső soron kizárólag nyereségre kell törekedniük ahhoz, hogy fennmaradjanak. Az információk szakmai személyes tudássá való átalakulása tudományos, közművelődési és kulturális keretek között valósul meg, s éppen az a keret, melyekben a könyvtárak működnek.

5. Keresés az interneten¹⁶

5.1 A web dokumentumai

A távoli online információkeresés – olykor már a helyi is – ma túlnyomórészt azzal kezdődik, hogy mind a laikus, mind a professzionális felhasználó a weben keresztül kapcsolatba lép az általa kiválasztott adatbázissal. Ez az a kályha, ahonnan mindenki elindul.

A weben keresztül azonban nemcsak az online elérhető adatbázisok váltak széles körben és felhasználóbarát keretek között elérhetővé, hanem a digitálisan tárolt információk végeláthatatlan állománya is. Ez új távlatokat s egyben nagy kihívást is jelent a könyvtári információkeresés számára.

A weben a kereső elsősorban a hipertextes adatátviteli szabványok (HTTP, Hypertext Transfer Protocol) szerint elérhető dokumentumokkal találkozik. Közöttük ma szinte kizárólagos szerepe van a speciális hipertext formátum (HTML, Hypertext Markup Language) szerinti szerkezetű dokumentumoknak, noha várható, hogy a jövőben a HTTP-n belül más (például XML) szerkezetű dokumentumok is fokozottan el fognak terjedni.

A HTML-dokumentumok egy vagy több weblapból (weboldalból) állhatnak.

A weblap fontos része az ún. címfej (header), melyben a legfontosabb másodlagos adatok – a HTML-dokumentum ismerveit – helyezhetők el. Ezeket a web körülményei között metaadatoknak nevezik.

A címfej metaadatai csak a HTML-dokumentum forrásnyelvén láthatók, külön utasítás nélkül nem jelennek meg a címlapon, viszont a keresőeszközök számára minden további nélkül elérhetők. Példaként az OSZK-tezaurusz weblapjának címfejét mutatjuk be a 19. ábrán.

```
<HTML>
<HEAD>
<TITLE>OSZK tezaurusz</TITLE>
<META NAME="author" CONTENT="Ungváry Rudolf">
<META NAME="description" CONTENT="Egyetemes osztályozásra alkalmas, átfogó
tezaurusz, melyhez taxonómiai és földrajzi, valamint történelemföldrajzi tezauruszok is tar-
toznak">
<META NAME="keywords" CONTENT="tezaurusz, OSZK, Országos Széchényi Könyvtár,
National Szechenyi Library, átfogó tárgykör, átfogó tezaurusz, egyetemes tezaurusz, uni-
verzális tezaurusz">
<META NAME="classification" CONTENT="025.43:001">
</HEAD>
```

19. ábra. Az OSZK-tezaurusz kezdőlapjának címfeje. A metaadatok ékes zárójel között szerepelnek és ismétlődhetnek. A META NAME után idézőjelek között látható a metaadat típusának a neve, a CONTENT után az értéke. A “description” az annotációt, a “keywords” a kulcsszavakat “(szabad tárgyszavakat”, deskriptorokat, jelzeteket), a “classification” az osztályozási jelzetet jelöli. Webhelye: <http://www.oszk.hu/ujdonsag/tezauroj.html>

A metaadat tehát szűkebb értelemben az internetforrások intellektuálisan vagy automatikusan előállított másodlagos adata, melyet HTML-dokumentumok esetén az elsődleges dokumentumba ágyaznak be. (A hagyományos dokumentumokban az – általában a címlap mögötti – impresszumban található a kiadásra, megjelenési helyre és időre stb. vonatkozó adatok.)

A metaadatok szabványosítása még nem érte el azt a fokot, ahol a bibliográfiai adatok szabványosítása tart.¹⁷ Ennél sokkal nagyobb hátrány, hogy a HTML-dokumentumok készítői

¹⁶ Szöveggyűjtemény, 2. köt. p. 430–528. Továbbá: Ungváry Rudolf: A tartalom szerinti információkeresés az interneten. I. és II. rész. Tudományos és Műszaki Tájékoztatás, 47. évf. 1. sz., 2000. jan. pp. 3–19; 47. évf. 3. sz., 2000. febr. pp. 55–67.

többnyire nem is élnek a címfej adta lehetőségekkel és nem határoznak meg metaadatokat. Legfeljebb a címet adják meg, de hiányzik a leírás (“description”) és a kulcsszavak (“keywords”). Igaz, hogy a készítőktől a szabványosítás jelenlegi fokán csak szabadon választott metaadatok megadása várható el, de még ezek az adatok is megbízhatóbb ismérvek lennének, mint a HTML-dokumentum szövegszavai, melyekre ma az interneten végzett kereséskor támaszkodni lehet. Emiatt a web dokumentumai manapság meglehetősen pontatlanul kereshetők.

Azt a weblapot, melynek afféle “címdoldal” vagy “kirakat” szerepet szánunk, nevezik **honlapnak** (home page, otthlap, hálószem). Ezen többnyire megjelenítenek valamiféle címet (az elektronikus dokumentumok leírói számára sokszor fogas kérdés, mi tekinthető egy ilyen lapon címnek), bevezető szöveget, de például a készítő, közreadó stb. adatai sokszor csak hosszas keresés után állapíthatók meg – ha egyáltalán megállapíthatók.

A HTML-dokumentum állhat egyetlen honlapból (szükségképpen egyetlen weblapból is), általában azonban több weblap kapcsolódik hozzá. A honlap legfontosabb elemei azok a HTML-kapcsolók, **csatolók** (az ún. “linkek”, “ugrópontok”), melyekre rákattintva a HTML-dokumentum többi weblapjára, továbbá más HTML-dokumentum weblapjaira vezetik tovább a keresőt. A keresésnek azt a módját, melyet a csatolók segítségével folytathatunk, nevezik “szörfölésnek” (surfing).

Azt a kezdőoldalt, melyről egy **intézmény** sokféle szolgáltatásának weblapjai elérhetők, **portálnak** nevezik (például a Hungary.Network Internetszolgáltató Rt. portálja, webhelye: <<http://www.net.hu/>>). Szerepük afféle “főbejárat” a felkínált információforrásokhoz. (Persze minden honlap eleve afféle portál, de az elnevezést ma a nagyobb kínálattal rendelkező intézményi honlapokra tartják fenn.) Ha ezeket a weblapokat hierarchikus rendszerbe szervezve kínálják, akkor **vortálról** (vertikális portálról) beszélnek (például az Axelero Internet Szolgáltató Rt. portálja, webhelye: <<http://www.axelero.hu/>>).

A weblapokat, a digitális (eredendően digitális formában készült) vagy digitalizált (hagyományosról digitálisra átalakított) dokumentumokat **elektronikus dokumentumoknak** nevezzük. Minden weblap digitális dokumentum, fordítva ez nem igaz: digitális vagy digitalizált dokumentumok még a mágneses (mágneslemezes, -szalagos) és optikai lemezes (CD-ROM, DVD-ROM stb.) hordozón tárolt dokumentumok is, melyek viszont megjelenhetnek weblapon is.

5.2 A web forrásai és forrásgyűjteményei

5.2.1 Adatbázisok és szöveges dokumentumok

A könyvtári katalógusok és az információkeresés egyéb forrásait képviselő adatbázisok az interneten a könyvtárak és más szolgáltatók honlapjain keresztül érhetők el:

- az egyes könyvtárak honlapjain, portáljain;
- könyvtárak csoportjainak részvételével működtetett közös katalógusok kezdőlapjain (például a különféle hazai könyvtári rendszerek együttes lekérdezését, az e rendszerek OPAC-jában egyidejű keresést lehetővé tevő KÖZELKAT, webhelye: <<http://www.kozelkat.iif.hu/>>, a Corvina rendszert használó könyvtárak “osztott” katalógusa, a VOCAL, webhelye: <<http://vocal.lib.klte.hu/>>) és az előkészületben lévő országos közös katalógus, a MOKKA, webhelye: <<http://www.mokka.hu/>>;
- azoknak a kereskedelmi szolgáltatóknak a honlapjain, melyek könyvtári szempontból releváns adatbázisokat tartanak fenn, (például a DIALOG adatbázis-szolgáltató, webhelye: <<http://www.dialog.com/>>). (A legfontosabb szolgáltatóknál elérhető adatbázisok típusait és fő használóikat a 4.2 fejezetben ismertettük.)

¹⁷ A legjelentősebb szabvány jellegű formátum a dublini alap(mag)-metaadatok formátuma, melynek 1999. évi változata 15 adatelemet tartalmaz. Lásd: Dublin Core Metadata Element Set. Version 1.1 Reference Description. Recommendation. <http://purf.org/dc/about/element-set.htm>

Létezik számos olyan szolgáltatás, melynek honlapja útján meghatározott szempontok szerint kiválogatott, interneten elérhető könyvtári weblapok, honlapok, portálok sora érhető el (nevezik az ilyen szolgáltatásokat linkgyűjteménynek is). Például:

- a weben elérhető nemzeti és más nagy könyvtárak a világban = Web Accessible National and Major Libraries (a könyvtári szövetségek nemzetközi csúcsszerve, az IFLA által fenntartott linkgyűjtemény), webhelye: <<http://www.ifla.org/II/natlibs.htm>>;
- az európai nemzeti könyvtárakhoz készült webhely, a Gabriel Gateway to Europe's National Libraries (a “gateway” jelentése “átjáró”, különböző kommunikációs szabványokat használó hálózatok összekötésére való eszköz), webhelye: <<http://www.k.nl/gabriel/en/>>;
- hungarika-könyvtárak a világban: Hungarika WWW. Külföldi könyvtári hungarika-gyűjtemények; webhelye: <<http://www.oszk.hu/hungarik/>>;
- Magyarország határain kívüli egyéb (tehát nem csak nemzeti) könyvtárak: LibWEB. Library Servers via WWW (a Berkeley Digital Library szolgáltatása); webhelye: <<http://sunsite.berkeley.edu/Libweb/>>;
- magyar könyvtárak: HUNOPAC. Könyvtári információ Magyarországon (integrált könyvtári rendszerek, adatbázisok, könyvtári sajtó, levelezőlisták); webhelye: <<http://www.mek.iif.hu/porta/virtual/magyar/opac.htm>>
- NIIF adatbázisok (a Nemzeti Információs Infrastruktúra Fejlesztési Program weben elérhető adatbázisai, például tankönyv adatbázis), webhelye: <http://www.iif.hu/db/>.

A könyvtári információkeresés számára hasznos ugrópontokat tartalmaznak a könyvtárszakmai portálok is:

- Könyvtárkapu. Információforrások könyvtárosoknak, webhelye: <http://www.bibl.u-szeged.hu/mke_eksz/portal/index.html>
- Könyvtárlap, webhelye: <<http://konyvtar.lap.hu/>>
- Könyvtár portál, webhelye: <<http://www.irodaweb.hu/konyv/konyvtar.htm>>

Ilyen dokumentumok más országok esetében is rendelkezésre állnak. Az Egyesült Államok szövetségi hatáskörű forrásairól például a Kongresszusi Könyvtár kézikönyvéből tájékozódhatunk:

- Handbook of Federal Librarianship, webhelye: <<http://lcweb.loc.gov/flicc/hndbk.html>>

A katalógusok és egyéb, dokumentumrekordokat tartalmazó adatbázisok csak töredékét alkotják a web eszközeivel elérhető forrásoknak (HTML-dokumentumoknak), noha ez a töredékhalmoz a legértékesebbek közé tartozik.

Az interneten elérhető dokumentumok összességéhez képest ugyan csak elenyészően kicsi, de az adatbázisokhoz hasonlóan fontos részét alkotják **a teljes szöveget tartalmazó digitális vagy digitalizált dokumentumok.**

Teljes szövegű szépirodalmi vagy tudományos dokumentumot bárki elhelyezhet az interneten. Az elektronikus publikációnak ugyanis előnye, hogy nincsenek nyomtatási költségei, rövid a megjelenés átfutási ideje és a világ minden táján könnyen, olcsón elérhetők, letölthetők, kinyomtathatók – feltéve, ha megtalálják őket. A publikációk kiszámíthatatlan szétszórtságban jelennek meg: személyi vagy intézményi honlapokhoz csatolva, vagy internetes – többnyire elektronikus időszaki – kiadványokban. További keresési problémát okoz, hogy a teljes szövegű elektronikus dokumentumok egy része hordozható dokumentumformátumban (PDF, Portable Document Format) kerül az internetre, tehát nem HTML-formátumban, s emiatt a csak HTML-re szakosodott keresőrendszerek nem találják meg őket. A keresési nehézségekre az “5.3.4 Láthatatlan web” című fejezetben még visszatérünk.

Meg kell jegyezni, hogy az interneten elhelyezett publikációk értéke és tekintélye összehasonlíthatatlanul kisebb, mint a hagyományos vagy rangos elektronikus folyóiratokban megjelent, előzetes lektori szűrőkön átesett közleményeké.

Az **elektronikus dokumentumok gyűjteményét** elektronikus könyvtárnak nevezik. Bennük (pontosabban a szervergépükön) a dokumentumokat eredeti, teljes szövegű formában tárolhatják, de

gyakoribb, hogy csak a leírásukat tárolják, azaz az adott könyvtár szempontjából ezek csak virtuálisan létező – csak hálózaton keresztül elérhető – dokumentumok. Például:

- Magyar Elektronikus Könyvtár (tartalmaz teljes szövegű dokumentumokat és leírásokat is); webhelye: <<http://www.mek.iif.hu>>;
- WebKat.hu, a Neumann János Digitális Könyvtár katalógusa, a magyar kulturális örökség körébe tartozó, interneten elérhető elektronikus dokumentumokról; webhelye: <<http://www.neumann-haz.hu/cgi-bin/webkat>>;
- National Library of Canada Electronic Collection; webhelye: <<http://collection.nlc-bnc.ca>>;
- Gallica; webhelye: <<http://gallica.bnf.fr>>.

Az interneten egyre több képi információ is található. Ide tartoznak a digitális térképek, fényképek, mozgóképek és animációk. Számos olyan gyűjtemény létezik, amelyben a képi dokumentumok valamilyen szöveges ismérv (például szerző, festő vagy téma) alapján kereshetők. Gyakorik a múzeumok, galériák műtárgyait bemutató weblapok. Néhány példa:

- képzőművészet Magyarországon a kezdetektől a XX. század közepéig; webhelye: <<http://www.kfki.hu/keptar>>;
- a párizsi Louvre gyűjteménye; webhelye: <<http://louvre.fr/>>;
- a firenzei Uffizi gyűjteménye; webhelye: <<http://www.televisual.net/uffizi/indice.html>>;
- a bécsi Kunsthistorisches Museum gyűjteménye; webhelye: <<http://www.khm.at/khm/staticE/page1.html>>;
- álló- és mozgóképek kereskedelmi forgalmazása: Getty Images Creative; webhelye: <<http://creative.gettyimages.com/>>;
- linkgyűjtemények, webhelye: <<http://www.icom.org.vlmp/>>; <<http://wwar.com>>.

Térképek linkgyűjteménye:

- történelmi, jelenkori és 3D térképek, irányítószámok a világ minden tájáról, webhelye: <<http://www.webmania.hu/>>;
- térképek a web magyar részén; webhelye: <<http://lazarus.elte.hu/hun/moterkep.htm>>.

Hangdokumentumok és letöltésükhöz szükséges programok, linkgyűjtemények:

- Zene tömörített formátumban, és amivel lejátszható vagy tömöríthető; webhelye: <<http://www.mp3.webmania.hu/>>
- The Internet's Music-Only Search Site, webhelye: <<http://www.musicsearch.com/>>
- a Yahoo! hangportálja, webhelye: <<http://de.music.yahoo.com/>>.

Az internetről letölthetők szoftverek is. Regisztrációs szoftvereket (“próbáld ki, mielőtt megveszed szoftver”, shareware) tartalmazó gyűjtemény:

- Yahoo! Shareware; webhelye: <<http://de.shareware.yahoo.com/>>.

5.2.2 Különleges HTML-dokumentumok

Lényegesen nagyobb azoknak a HTML-dokumentumoknak a száma, melyek nem sorolhatók az előbbi fejezetben tárgyalt típusokhoz, mégis fontos források lehetnek az információkeresők számára.

Egy részük a hagyományos dokumentumok elektronikus változata (elektronikus könyvek, tanulmányok, folyóiratok).

Más részük bibliográfiai szempontból keményebb dió¹⁸. A könyvtári gyakorlatban a vállalati dokumentumokat (használati utasítások, jelentések, tervek), igazgatási dokumentumokat (közlemények, előírások), oktatási dokumentumokat (tanrendek, módszertanok), továbbá a plakátokat és az aprónyomtatványokat (röplapok, színlapok, brosúrák, prospektusok, címkék,

¹⁸ Sipos Márta, Ungváry Rudolf: Hivatkozás távoli hozzáférésű HTML-dokumentumokra. In: Tudományos és Műszaki Tájékoztatás, 47. évf. 12. sz. 2000. dec. pp. 495–502.

jegyek, papírpénzek, részvények, halotti jelentések stb.), menetrendeket, telefonkönyveket stb. “nem hagyományos dokumentumoknak”, “vállalati irodalomnak”, “szürke irodalomnak” vagy összefoglalóan használati dokumentumoknak (ami semmitmondó, félrevezető név, mert nem csak ezek a dokumentumok “használatiak”) nevezik. Ezek a dokumentumok mostohagyuermekei a bibliográfiáknak és a katalogizálásnak. Az internet dokumentumainak jelentős része használati dokumentumnak tekinthető. Túlnyomórészt testületek hozzák létre, de magánszemélyektől is származhatnak.

Használati dokumentumoknak tekinthetők a testületi honlapok is (hasonlítanak a hagyományos vállalati dokumentumokhoz), részben a röplapokra, a plakátokra, idegenforgalmi prospektusokra, hirdetésekre, áruajánlatokra emlékeztetnek, nagyon gyakran részletes termékinformációkat adnak. Ugyancsak használati dokumentumoknak tekinthetők a személyi honlapok. Ezek afféle részletes névjegykártyák, melyek a személyi adatoktól kezdve az életrajzon át az érdeklődési területekig mindenfélét tartalmazhatnak.

Mindezeknek a HTML-dokumentumoknak a lehető legváltozatosabb a tartalma, és ami az impresszumukat (megjelenési adataik összefoglaló közlését) illeti, a lehető legszegényesebbek. Leírásuk komoly nehézségeket okozhat,¹⁹ különösen, ha nem tudjuk, hogy az esetek egy részében a szerzőségi és megjelenési adatokat megadják a HTML-dokumentum címfejében (és csak ott).

Az alábbiakban megkísérlünk áttekinthető, de csak közelítő tipológiát adni a felsorolt weblapokról.

Tudományos és ismeretterjesztő HTML-dokumentumok:

- elektronikus könyvek és folyóiratok;
- tanulmányok (sokszor PDF-formátumban), oktatási (tankönyv/jegyzet jellegű) weblapok, felvilágosító szövegek.

Kereskedelmi HTML-dokumentumok:

- vállalati és vállalkozói honlapok (termékinformációk, árukatalógusok, szolgáltatások);
- tőzsdei és banki szolgáltatások;
- internet üzletek, áruházak, “HTML-prospektusok”, ajánlatok;
- hirdetések (jegyirodák, utazási irodák stb., állásajánlók).

Politikai és kulturális HTML-dokumentumok:

- politikai célzatú HTML-dokumentumok (például politikai hirdetések, felhívások, pártok, mozgalmak honlapjai);
- kormányzati, igazgatási HTML-dokumentumok (például szervezeti/ügyviteli ismertető, pályázati kiírások);
- kulturális és vallási célzatú HTML-dokumentumok (például ezoterikus szövegek, imák, egyházak, vallások honlapjai);
- szórakoztató és turisztikai célú HTML-dokumentumok (sportoldalak, túraleírások, webkamerák, webkávézók, csevegők, társskereső oldalak, játékok, viccek, horoszkópok, szexoldalak stb.).

Hírszolgáltatások HTML-dokumentumai:

- politikai, gazdasági/tőzsdei és sporthírek;
- aktualitások, időjárás, vízállás;
- műsorok, rendezvények programjai;
- személykeresés, rendőrségi és egyéb felhívások;
- eseménynaptárak.

Személyi honlapok.

Mindezekre a dokumentumokra általában jellemző, hogy

- többnyire rendkívül hiányosan az azonosító adataik (címük, közreadójuk, készítőjük, stb.);
- időszakiságuk, megjelenési/frissítési gyakoriságuk sokszor megállapíthatatlan;

¹⁹ Az elektronikus dokumentumok bibliográfiai leírását az ISBD(ER) szabványon alapuló magyar KSZ 2 szabályozza, a rájuk való hivatkozást pedig az ISO/DIS 690–2 szabvány <<http://ifla.inist.fr/VII>> <http://www.ifla.org/l/training/citatio/citing.htm>.

- ha részdokumentumok, nem mindig könnyű behatárolni, melyik a gazdadokumentum, olykor nincs semmi összefüggés a gazdadokumentummal (más szóval nincs feltétlenül tartalmi összefüggés honlap és hozzá csatolt weblap között).

5.3 A web keresőeszközei és keresőszolgáltatásai

5.3.1 A keresőszolgáltatások fajtái

A HTML-dokumentumokhoz való gyors hozzáférést egyrészt az ún. kliensoldali böngészők, másrészt a szerveroldali keresőszolgáltatások biztosítják. A böngészők (browser, például Internet Explorer, Netscape) a felhasználó gépén működnek (ezeket azon “futtatják”), a keresőszolgáltatások pedig a szolgáltatók szervergépein. A böngészők ún. ügyfélalkalmazások, melyek lehetővé teszik a felhasználó számára, hogy HTML-dokumentumokat nézzen meg a weben, más hálózaton vagy a saját számítógépén, kövesse közöttük a csatolókkal megvalósított hipertextes kapcsolatokat, és állományokat küldjön. A böngészőkkel tehát nem magát a keresést végzik.

A keresőszolgáltatások egy része globális, azaz – legalábbis elvileg – a teljes web a gyűjtőkörük. Másik részük nemzeti (egy állam doménneveire korlátozott), regionális (például az angol nyelvterület doménneveire korlátozott), vagy meghatározott nyelvre korlátozza gyűjtőkörét. Egy részük egyetemes, azaz mindenfajta és mindenféle tartalmú HTML-dokumentum a “gyűjtőkörükbe” tartozik, másik részük csak speciális tartalmú vagy speciális típusú HTML dokumentumokat gyűjt. Egyes szolgáltatások a gyorsaságukkal, mások a keresési eszközök gazdagságával, megint mások az adatbázisuk nagyságával tűnnek ki. Vannak a relevancia szempontjából megbízhatóbbak és kevésbé megbízhatóak. A weben rendszeresen megjelennek személyes kezdeményezésből készített elemzések, melyekben az egyes szolgáltatásokat minősítik.

A keresőszolgáltatásoknak két típusa alakult ki: az **indexelő szolgáltatások** és az **internetkatalógusok**. Az előbbieken belül különleges változatot képviselnek a **gyűjtő- és a metakeresők** (többszörösen indexelő szolgáltatások). A gyűjtőkeresőkben a keresőszolgáltatások választékát kínálják föl, egy szolgáltatást kiválasztva elvégezhető a keresés. A metakeresőkben egyszerre több keresőszolgáltatás is kijelölhető, és a rendszer mindegyik keresőszolgáltatás adatbázisában elvégzi a keresést, feltüntetve a találatoknál, hogy melyik szolgáltatás adatbázisában sikerült megtalálni a rekordot. (Az indexelő szolgáltatások és az internetkatalógusok magyarázatát a következő két fejezet tartalmazza.)

Az 5. táblázatban néhány ismertebb szolgáltatás látható.

	szolgáltatás	fő tartományi terület	nyelv	szakterület	
AltaVista	indexelő és katalógus	magyar	egynyelvű	általános	http://www.vindex.hu/ katalogus
Avatar Search – Search Engine of the Occult Internet!	katalógus	regionális	egynyelvű	esoterika	http://www.avatarsearch.com/
EuroDocs: Primary Historical Documents	katalógus	regionális	egynyelvű	európai történelem	http://www.lib.byu.edu/~rdh/eurodocs
Google	indexelő	globális	többnyelvű (magyar is)	általános	http://www.google.com/
Heureka	metakereső	magyar	egynyelvű	egyetemes	http://www.heureka.hu
HuDir	katalógus	magyar	egynyelvű	általános	http://www.hudir.hu
MetaGer	metakereső	globális	egynyelvű	általános	http://meta.rzn.uni-hannover.de
Northern Light	indexelő és automatikusan osztályozó	globális	egynyelvű	általános	http://www.northernlight.com
PsychCrawler	indexelő	regionális	egynyelvű	pszichológia	http://www.psychcrawler.com
Yahool	katalógus	globális	többnyelvű	általános	http://www.yahoo.com

5. táblázat. Néhány ismertebb keresőszolgáltatás főbb jellemzői. A magyar és magyar nyelven is használható rendszerek neve félkövéren szerepel. Aláhúzással a globális külföldi példákat jelöltük (az aláhúzatlan speciális rendszereket csak a választék érzékeltetésére szerepeltettük).

Ahhoz, hogy a rendelkezésre álló keresőszolgáltatásokat gyorsan át lehessen tekinteni és kiválasztani, a professzionális könyvtári információkeresőnek (köztük a könyvtári referenzszolgálat tagjainak is) célszerű ismernie egyrészt

- a könyvtári tárgyú portálokat (például Könyvtárkapu, Könyvtárlap, Könyvtár Portál), ahonnan a fontosabb szolgáltatásokat a csatolókon keresztül elérheti; nem utolsósorban pedig
- azokat a jelentősebb linkgyűjteményeket, melyek a különféle keresőszolgáltatások választékát tartalmazzák (többek között tematikusan is rendezve).

Nagyon sok szakterületi keresőszolgálat létezik, és minél speciálisabb egy internetes keresőszolgálat, általában annál megbízhatóbbak a segítségével nyerhető találatok. A korszerű referenzszolgálat számára követelmény, hogy ezeknek a szolgáltatásoknak az alapján tájékoztatni tudjon, illetve az olvasóknak a használatukat javasolja.

Vannak olyan honlapok, melyek arra szakosodtak, hogy meghatározott témákban felkínálják a fontosabb weblapokat. A weblapok, szolgáltatások nevei egyben csatolók (ugrópontok, linkek), rájuk kattintva elérhető a weblap. Az ilyen linkgyűjtemények – mint a weblapok általában – a hagyományos dokumentumoknál sokkal jobban ki vannak téve a változásoknak, avulásnak. Ami ma létezik, holnap már nincs. A professzionális keresőnek tehát folyamatosan “jelen kell lennie” az ún. “kibertérben”, hogy tisztában lehessen az aktuálisan használható keresési segédeszközökkel. Néhány jelenleg létező linkgyűjtemény (a választék ennél lényegesen nagyobb):

- Keresők gyűjteménye. Hungary.Network;
webhelye: <<http://www.net.hu/search>>
 - Search Engines Showdown. The Users' Guide to Web Searching;
webhelye: <<http://www.bubl.ac.uk/link/>>
 - Suchdienste im Internet – Search & Find in the Internet;
webhelye: <http://www.bsz-bw.de/links/suchdienste.html>
 - WorldWide directory for Webmasters²⁰;
webhelye: <<http://weblinea.netlinea.com/en/>>
 - Selected Internet resources covering all academic subject areas, webhelye: <<http://www.bubl.ac.uk/link/>>
 - Különleges keresők tematikusan csoportosított gyűjteménye Chris Sherman “The Invisible Web” c. könyvéből;
webhelye: <<http://www.invisible-web.net/>>
- Végül, de nem utolsó sorban, a felkészült keresőnek ismernie kell az olyan fontosabb honlapokat és portálokat, melyekben információk találhatóak a közigazgatásról, a közintézményekről stb. Például:
- Magyarország honlapja (megyék, városok, intézmények); webhelye: http://www.fsz.bme.hu/hungary/homepage_h.html

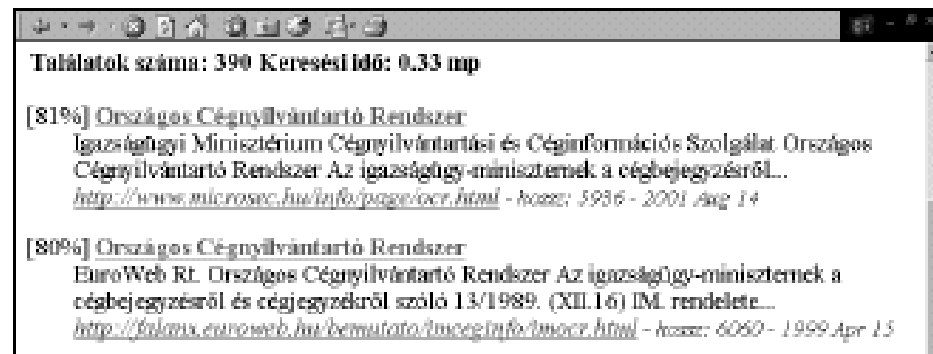
5.3.2 Indexelőszolgáltatások (“keresőgépek”)

A indexelőszolgáltatások azon alapulnak, hogy az ún. “keresőgépek” (valójában keresőprogramok) rendszeresen pásztázzák (végig-”szörfölik”) a hálózatot a dokumentumok csatolói (linkjein) keresztül, megállapítják a HTML-dokumentum legfontosabb – és egyáltalán: megállapítható – adatait, és a dokumentum szövege alapján a tartalmi ismérveit (az indexkifejezéseket). Többségük nem a teljes szöveget nézi át az indexelés érdekében, hanem csak a dokumentum elején meghatározott számú (például 20–30) sort, és egy részük figyelembe veszi a címfejlben szereplő metaadatokat is – ha léteznek. A szolgáltatások általában arra törekszenek, hogy a teljesség legyen

²⁰ A “Webmasterss” nem sajtóhiba, hanem így szerepel

nagy, ezért pontosságról eleve nincsen szó (lásd a “8. Az információkeresés értékelése” c. fejezetet). Általában relevancia-ellenőrzést is végeznek, és a találatokat ennek sorrendjében jelenítik meg. (A relevancia általuk meghatározott mértéke többnyire attól függ, hogy milyen gyakran fordul elő egy szó a szövegben, előfordul-e együttesen a címben, a címfejben és a szövegben is, csatolóként ugrópontként kijelölték-e a szót a szövegben belül stb.)

A HTML-dokumentumok azonosító adataiból álló HTML-rekordokat adatbázisba gyűjtik, a rekordok az indexkifejezések alapján kereshetők. A HTML-rekordok az indexelt dokumentumok leírásait tartalmazzák. A 20. ábrán ilyen rekord megjelenítése látható találatként.



20. ábra A Heureka indexelőszolgáltatás megjelenített HTML-rekordja. Keresés a “Cégnyelvántartás” keresőszóval. A legfelső sor baloldalán a relevancia mértéke [81%], utána a cím, alatta a dokumentum szövegének kezdete (ahogy a cím alatt a szöveg kezdődik), alatta a dokumentum azonosítója, utána a hossza és utolsó frissítésének dátuma látható. Webhelye: <<http://www.heureka.hu>>

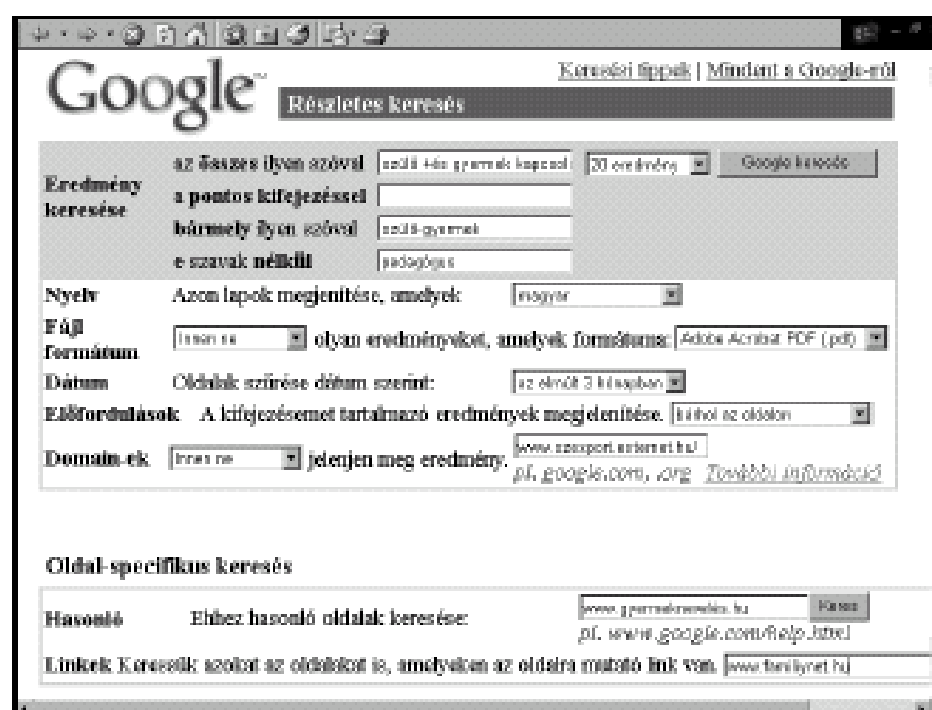
Az alkalmazható keresési stratégiák a keresőrendszerrel függenek és meglehetősen változatosak. Az adatbázisokban alkalmazható paranccsal vezérelt kereséshez képes (egyelőre) valamivel szegényesebb a választék, de ez a helyzet változóban van: a keresőszolgáltatók már ma is számos olyan keresési lehetőséget biztosítanak, melyekre a hagyományos, paranccsal vezérelt kereséskor nincs lehetőség.

Néhány keresőszolgáltatásban (AltaVizsla, Northern Light) automatikus osztályozás támogatja a keresőt abban, hogy a kérdésének megfelelő találatok mellett még a hasonló találatok is megjelenjenek. Más rendszerekben a találat azonosítójához (az URL-hez) hasonló azonosítójú rekordokat kínálnak föl automatikusan. Ezeket a lehetőségeket sokszor a “powered search” (“megerősített”, “megtámogatott” keresés) kifejezéssel jelzik.

Alapértelmezésben a keresőrendszereket a legtájékoztatlanabb felhasználó számára alakították ki. Ha több keresőszót adunk meg szóközzel elválasztva, a rendszer automatikusan ÉS–kapcsolatnak értelmezi.

Minden rendszerben található részletes tájékoztatás a keresési lehetőségekről, ezekhez általában több beviteli mezőből álló keresőfelületeket kínálnak.

Mivel az indexelt HTML-dokumentumok állománya teljesen ellenőrizetlen, az egyszerű kereséskor rendkívül nagy a felesleges találatok száma, a zaj. Mindenképpen érdemes kihasználni a részletes (összetett, “haladó”) keresési lehetőségeket (advanced search, powered search). A 21. ábrán egy indexelő szolgáltatás részletes keresést biztosító felülete látható.

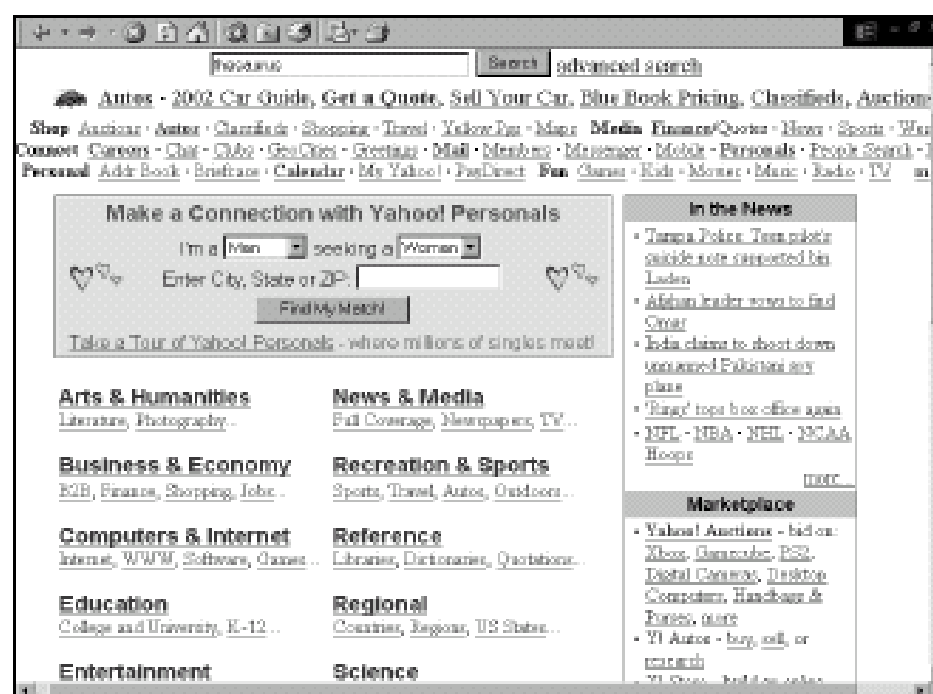


21. ábra A Google felülettel vezérelt keresési űrlapja. A jobb felső sarokban a keresési tippek, lent jobbra pedig a További információ tájékoztatnak a részletes keresési lehetőségekről. “Szülők és gyermekek iskolán kívüli kapcsolatai” volt a kérdés, kizárva a szexoldalakat, a PDF (hordozható formában kódolt) dokumentumokat és a három hónapnál régebbi weblapokat. Az “Összes ilyen szóval” mezőben az “és előtti + jel azt jelenti, hogy a kereső ne tiltott szóként kezelje az “és” szót. A Google webhelye: http://www.google.com/advanced_search?hl=hu

5.3.3 Internetkatalógusok (“directory”-k)

Az internetkatalógusok (directories, browsing services) **hierarchikus osztályozási rendszert alkalmazó keresőszolgáltatások**. Adatbázisaik többnyire intellektuálisan feldolgozott HTML-dokumentumok rekordjait tartalmazzák, valamint kapcsolatokat más adatbázisokhoz. Az osztályozást és a tartalmi kivonatok megfogalmazását szerkesztőségben végzik, de a weblapok fenntartói maguk is kérhetik weblapjaik fölvételét a katalógusba, s ilyenkor ők határozhatják meg a weblap besorolását és ők szövegezhetik meg a rövid tartalmi kivonatot.

A katalógusokban az osztályok alapján – “lapozgatva” – végezhető böngészés, de lehetőség van arra is, hogy a keresőszó megadásával, célzott kereséssel érjünk el a kívánt osztályig. A 22. ábrán egy internetkatalógus kezdőlapja látható.



22. *ábra* A Yahoo! internetkatalógus portálja. A képmező közepétől lefelé balra a hierarchikus osztályozási rendszer legfelső szintjének osztályai láthatók betűrendben. Fönt négy sorban más adatbázisokhoz (keresőkhöz) lehet hozzáférni. Jobbra látható a hírhasáb. Legfelül a “search” mezőben adható meg a keresőszó, melynek alapján a rendszer a megfelelő speciális osztályt jeleníti majd meg, ha a “search” parancsot megadják. A képen nem látható alsó sorokban a különféle nyelvű Yahoo! változatok és egész sor speciális *adatbázis (keresőszolgáltatás) ugrópontjai található.* Webhelye: <<http://www.yahoo.com>>

Léteznek olyan katalógusok is, melyek indexelőszolgáltatásként is működnek (ilyen például a AltaVizsla). A keresőszót megadva egyrészt megjelennek az indexelőszolgáltatás segítségével nyert találatok, másrészt a rendszer fölkinálja a releváns osztályokat is. Ritkán tezaurusz is kiegészíti az osztályozási rendszert (példa ugyancsak az AltaVizsla). A tezaurusz lexikai egységeit az AltaVizslában összekapcsolták a katalógus osztályaival is; bármelyik deszkriptort megadva ugyancsak lejátszódik a keresés az indexelőszolgáltatás segítségével, ugyanakkor azoknak az osztályoknak a találatai is megjeleníthetők, melyekre a deszkriptor mint mutatószó utal.

Az internetkatalógusok adatbázisai lényegesen kisebbek, mint az indexelőszolgáltatásokéi, de a keresés az intellektuális feldolgozás, a gondos osztályozás jóvoltából sokkal kisebb zajjal jár. A megjelenített tételek jobban tájékoztatnak a dokumentum tartalmáról, mert a tételekhez a feldolgozók vagy a beküldők (akiknek érdekében áll, hogy a katalógusba bekerüljön a weblapjuk) szövegezik meg a tartalmi kivonatot. Ahhoz képest, hogy egy indexelőszolgáltatásban mi mindent kell ismerni a pontos részletes kereséshez, az internetkatalógusokban sokkal egyszerűbb az eljárás, noha kevesebb a találat.

A szakterületekre specializálódott keresőszolgáltatások többsége internetkatalógus. Egy részüket neves kutatóintézetek gondozzák, más részük kereskedelmi szolgáltatás. Mindkét esetben jó minőségű találatokra lehet számítani. Az osztályozási rendszerek között nemcsak az általánosan elterjedt, természetes nyelven alapuló hierarchikus osztályozási rendszerek fordulhatnak elő, hanem az ETO (az Egyesült Államokban pedig a Dewey-féle TO) is.

5.3.4 A “láthatatlan web”

A HTML-dokumentumokra is érvényes, hogy az elektronikus információforrások nem helyettesítik a nyomtatott forrásokat, hanem kiegészítik. Nem igaz, hogy az interneten minden információ megtalálható. A weblapok látványosak, információik könnyen emészthetőek, de mögöttük csak ritkán van mélyreható tartalom. Válogatás nélkül kerülnek az internetre, az indexelőszolgáltatások keresőprogramjai csak formális relevancia-ellenőrzést végezhetnek, az internetkatalógusokban pedig legfeljebb a weblapok töredékét sikerül intellektuálisan feldolgozni.

Nem csak az a probléma az internettel, hogy a célzott kereséskor is rendkívül nagy a zaj, hanem az is, hogy az indexelőszolgáltatások keresőprogramjai sem képesek elérni minden dokumentumot. A sokféle ok közül néhány fontosabb:

- a keresőprogramok nem mindegyike szörföl végig minden tartományt (doménnevet); az alkalmazott “begyűjtési stratégia” (gathering, harvesting) és forrásfelkutatás (resource discovery) függvénye, hogy milyen szervereket talál meg a program, és azon belül milyen dokumentumok indexelését részesíti előnyben;
- szolgáltatásonként változik, hogy mi számít indexelendő egységnek a program számára; van, amelyik minden elért dokumentumot számításba vesz, de csak a töredékét indexeli, megint mások csak a teljes szövegükben indexelt dokumentumokat veszik figyelembe stb.;
- az “először átfogóan” indexelő stratégiát alkalmazó rendszerek kevésbé részletesen, de sok dokumentumot indexelnek, az “először mélyen” indexelő stratégia eredménye pedig a részletesen indexelt, de kevés dokumentum lesz;

- az aktualizálás (frissítés) gyakorisága a hetenkénti frissítéstől a félévenkénti, sőt évenkénti frissítésig terjedhet; minél nagyobb a szolgáltatás, annál kisebb frissítési gyakoriságra számíthatunk (az ún. halott tételek száma egyes nagy rendszerben 20–30% is lehet);
- az interneten jelentős számban található nem HTML-formátumú – hanem például PDF-formátumú – dokumentumok, melyek indexelésére nem minden szolgáltatás képes (jóllehet a Google például lehetővé teszi akár a Word-formátumú dokumentumok keresését is), holott éppen a PDF-formátumú dokumentumok között található a legértékesebb tanulmányok és egyéb szöveges dokumentumok;
- számos weblap elejének szövege (melyet az indexelő programok átnéznek) alkalmatlan arra, hogy automatikus indexeléssel a tartalmát megfelelően reprezentálja.

Mindezek következtében az internetnek létezik egy láthatatlan része (invisible web). Egyes szerzők szerint az Internet több mint 80%-át a globális indexelőszolgáltatások keresőprogramjai nem érik el. Sherman szerint²¹ ide tartoznak a nyilvános online katalógusok (the “opaque” web), a személyi honlapok (the private web), az amatőrök által az interneten elhelyezett dokumentumok (the proprietary web) és végül a “valódi” láthatatlan honlapok (the truly invisible web). Az utóbbiak közé tartoznak a PDF, a Word stb. formátumú dokumentumok.

Noha mindebben van némi túlzás annak érdekében, hogy a web problémáival foglalkozó dokumentumok keresletét növeljék, tény, hogy számos esetben nem hagyatkozhatunk pusztán az indexelőszolgáltatásokra és internetkatalógusokra.

A legegyszerűbb esetek közé tartozik, amikor valóban különleges (például PDF-, WORD-) formátumú dokumentumokra kellene ráakadnunk. A keresőszolgáltatások tulajdonságainak mélyebb ismeretében ezen a problémán adott esetben segíteni lehet.

A könyvtárak nyilvános online katalógusai a könyvtárak honlapjainak ismeretében többnyire elérhetők, illetve tájékozódni lehet arról, hogy hogyan lehet távoli hozzáféréssel másképpen belépni a katalógusokba, ha az interneten keresztül ez nem lehetséges.

Olykor az is segít, ha – szokatlan módon – megpróbáljuk kitalálni a weblap azonosítóját, az URL-t. Ha például a “szemnek láthatatlan jelenségekről” akarunk tájékozódni, az is lehetséges, hogy ez már mást is érdekelt, és létezik a www.láthatatlan.hu vagy a www.rejtett.hu stb. azonosítójú weblap. Vannak keresők (például a MetaGer, lásd az 5. táblázatot), melyben gyorsan megállapítható, létezik-e egyáltalán a keresett URL. A Google vagy az AltaVizsla segítségével pedig kereshetünk hasonló azonosítójú weblapokat.

Vannak azonban nehezebb esetek. Például olyankor, ha nem csak nagyon speciális kérdésben akarunk tájékozódni, de ráadásul nagyon új vagy nagyon különleges problémáról van szó, melynek sokféle nyelvi megfogalmazása lehet. Ehhez hasonló, de még “súlyosabb” a helyzet, ha nem is tudjuk pontosan, mit akarunk keresni. Ez utóbbi eset előfordulása nagyon is valószínű, és egyáltalán nem az ostobasággal, hanem az alkotómunka természetével függ össze.

Amikor még megoldatlan tudományos kérdésekről van szó, vagy arról, hogy eredményeket kell összegezni, értékelni, nem lehet mindig előre pontosan megmondani, mire is akarunk vagy mire kell kilyukadni. Ez csak sejthető.

A megoldás pontos megfogalmazására ugyanis éppen a problémamegoldás során lehet csak rájönni. Ezért a megoldást előre nem is lehet pontosan keresni.

Ilyenkor inkább a kissé határozatlan, inkább az ösztönök, az intuíció által vezérelt bóklászás, többek között a hagyományos dokumentumokban való lapozgatás és a weblapok közötti kíváncsi ide-oda kerességélés segít. Ez utóbbit a csatolók segítségével végezzük. A szövegben elhelyezett ugrópontokra kattintva más dokumentumokat keresünk föl, és azt várjuk, hogy ezek az “ugrások”, mint valamiféle asszociációk önkéntelenül is rávezessenek arra, amit valójában keresünk. A szörfölésnek ezt a módját nevezik innovatív keresésnek, találóbb angol kifejezéssel “serendipitous

²¹ Chris Sherman and Gary Price: *The Invisible Web. Uncovering Information Sources Search Engines Can't See.* – Los Angeles: Independent Publishers Group, 2001. 300 p. – (CyberAge Books) ISBN 0-910965-51-X

searching” a neve. “Szirének hangjára figyelő keresés...” Mindez már átvezet a művészet világába, mely nélkül azonban nem létezne se tudomány, se szakismeret.

6. Keresés a dokumentumok nem tartalmi jellemzői szerint

6.1 Általános jellemzők

Amikor az alábbiakban a “nem tartalmi” jellemzők szerinti információkeresésről lesz szó, valójában nem az információkereséssel, annak módszereivel és eszközeivel foglalkozunk (bár indokolt esetben erről is említést teszünk), hanem olyan dokumentumleírások készítésével vagy kereshető adatok alkalmazásával, amelyek nagy biztonsággal lehetővé teszik a dokumentumok keresését e nem tartalmi jellemzők alapján²². A “nem tartalmi jellemzők” valójában olyan jellemzők, amelyek nem vagy legalábbis az esetek többségében nem a dokumentum tárgykörét jellemzik, vagy ha mégis arra vonatkoznak, általában akkor sem készül belőlük előre kötött szótár, hanem az egységesítés érdekében szabványokba foglalt szabályokat állítanak fel. Ez utóbbira – az eleve egységes azonosító számok és jelek (kódok) kivételével – mindenkor szükség van, legalábbis elvben.

Az, aki információt keres nem tartalmi jellemzők alapján, az esetek többségében egy vagy több meghatározott dokumentumot (vagy – ha úgy tetszik – egy vagy több meghatározott információs egységet) keres, legyenek ezek egy szerző művei vagy meghatározott mű(vek), (kiadványok, dokumentumok) illetve azok változatai, egy kiadó kiadványai, meghatározott országban kiadott vagy meghatározott nyelvű dokumentumok stb. Mindezek a **keresési igények viszonylag egyszerűen kielégíthetők, ha a dokumentumok keresőképeinek alapjául szolgáló leírások úgy készülnek, hogy eleget tegyenek a fenti vagy hasonló szempontok szerinti keresés követelményeinek**. Ebből következik, hogy az ilyen ismérvek szerinti információkereséshez elég az olyan szabályrendszer és – esetenként – kódszavak (például országnévkódok vagy nyelvkódok) egyszerű szótára, amelyek segítségével e követelményeknek eleget lehet tenni. Az esetek jelentős részében e követelményeket azon szabványok betartása, alkalmazása biztosítja, amelyek a bibliográfiai leírás szabályait és a nem tárgyköri besorolási adatok egységesítésének előírásait tartalmazzák.

A kódrendszer – pontosabban kódszavak szótára – eleve kötött szótár. Például az országnevek vagy nyelvek kódszavai kötött szótárt alkotnak.

Még annyit kell bevezetesként elmondani a keresés alapjául szolgáló nem tartalmi jellemzőkről, hogy ezek sokkal kevésbé “rakoncátlanok” (ugyanis sokkal egyértelműbbek, pontosabban azonosítók) mint a tartalmi, mindenekelőtt tárgyköri jellemzők. A nem tartalmi jellemzők vagy eleve egységesek, egyedülállóak, összetéveszthetetlenek, vagy kisebb-nagyobb nehézségek árán azzá tehetők.

6.2 Nemzetközi (és más) azonosító számok és kódok

Bármennyire is az olvasók, következésképpen a könyvtárak szükséglete volt, hogy az egyes megjelent dokumentumokat vagy éppenséggel azok meghatározott kiadását egyértelműen kereshetővé tegyék, az erre szolgáló eljárások kezdetei a **könyvkereskedelemben születtek meg**. A széles körben terjesztett dokumentumok raktári nyilvántartását, megrendelését, szállítását általában is, a számítógépes módszerek elterjedése után pedig különösen, hasonlíthatatlanul

²² Meg kell említeni, hogy – akár csak a formai, illetve tartalmi feltárás esetében (1.3 fejezet) – az esetek jelentős részében nem arról van szó, hogy az alábbiakban tárgyalandó “nem tartalmi” jellemzőknek ne lenne tartalmi vonatkozása. Nem szorul bizonyításra, hogy van tartalmi jelentősége a szerzői vagy más funkciójú személyek, vagy testületek nevének, vagy a keresett dokumentum leírásában előforduló földrajzi neveknek, de még az azonosító számok/kódok esetén is találhatunk egyes tartalmi vonatkozásokat.

egyszerűbben, gyorsabban és megbízhatóbban lehetett intézni, ha rendelkezésre állt az adott dokumentum valamilyen egyértelmű, szabványos azonosító adata (száma). Mivel e számozás iránti igény a könyvkereskedelemből származott, az első azonosító szám nem művet azonosított, még csak nem is a mű adott kiadását, hanem adott kiadásának azonos kivitelű (fűzött, kötött stb.) és így azonos árú példányait. Az azonosító számnak egyértelműnek kellett lennie, így sem a kiadó, sem a terjesztő, sem a vevő nem engedhette meg magának az elírás fényűzését. Ezért az azonosító szám(ok)hoz a többi számjegyből egyértelmű algoritmus szerint kiszámítható ellenőrző szám járult. **Az első szabványos azonosító számot az SBN (Standard Book Number)** az Egyesült Államokban hozták létre, majd nemzetközi használatának terjedését nemzetközi szervezetének létrehozása, technikai megújítása, valamint országok (régiók, nyelvterületek) azonosítóinak felvétele a kiadót és a könyvet eleve azonosító számba tette lehetővé. Így jött létre az **International Standard Book Number (ISBN), vagyis a könyvek nemzetközi szabványos (azonosító) száma**, illetve annak használata, nemzetközi és soknemzeti szabványosítása. Nem sok időnek kellett eltelnie, hogy először – az UNESCO bábáskodásával – a gyakran meglehetősen rakoncátlan életutat befutó időszaki kiadványok (és ezen keresztül az ezekben megjelent cikkek azonosításának egy feltétele), majd más dokumentumfajták nemzetközi azonosító száma is létrejöjjen. Újabb mérföldkőnek tekinthetjük az internet és az elektronikus dokumentumok megjelenése és elterjedése nyomán létrejött újabb “azonosító család” kialakulását.

Ide kívánczik, hogy kezdetben (az ISBN hőskorában) a **könyvtárosok némi fanyalgással fogadták** az ISBN-t, különösen pedig azt, hogy az beépüljön a kor sok évszázados hagyományokra visszatekintő azonosító/kereső eszközeibe, a leíró katalógusokba, vagy hogy éppenséggel ISBN szerint elrendezett katalógusok épüljenek. A tartózkodás oka – eltekintve az újjal szemben szinte mindig megjelenő ellenállástól – az a fentiekben már említett tény volt, hogy az ISBN nem művet (bibliográfiai egységet), még csak nem is kiadást azonosított, hanem adott kivitelű/árú kiadást. Kétségtelen, hogy az azonosító számok katalógizálási célú, az ismert azonosító számú dokumentum keresését nagymértékben felgyorsító **használata a számítógépes, online használható katalógusokkal terjedt el általánosan.**

A nemzetközi azonosító számok adott dokumentumok részére történő kiadását és nyilvántartását nemzeti központok végzik – gyakran, így Magyarországon is – a nemzeti könyvtár keretében.

Mindeddig a nemzetközi azonosító számokról beszéltünk. Azonosító jelzete azonban nem egy esetben van **egy-egy intézmény vagy intézménycsalád által közreadott kiadványoknak** vagy hivatalos, kormányzati kiadványoknak is. Ezekről bővebben nem szólunk az alábbiakban, ami nem jelenti azt, hogy – az információforrásról szerzett értesülés tartalmától is függően – ezek nem tehetnének (például az ENSZ és szakosított szervezeteinek kiadványai, valamint más nemzetközi szervezetek kiadványai, hivatalos nemzeti kiadványok és különféle szervezetek mint kutatóintézetek, vállalatok és más intézmények kiadványai esetén) igen hasznos szolgálatot az információkeresőnek. Ez a könyvtárosoknak azt jelenti, hogy a bibliográfiai tételek készítésekor ne feledkezzenek meg az ilyen azonosítók leírásáról, illetve kereshetővé tételéről (bár természetesen már korábban, az állománygyarapítás során is hasznukat veszik). Kétségtelen azonban, hogy a legnagyobb figyelemre a nemzetközi azonosítók számíthatnak.

Ma “üzemszerű” használatban vagy közvetlenül bevezetés előtt – első lépésben eltekintve az interneten hozzáférhető elektronikus dokumentumok speciális azonosító számaitól – az alább felsorolt nemzetközi azonosító számok vannak:

Magyar (MSZ ISO) szabványok

MSZ ISO 2108:1994 Információ és dokumentáció. Könyvek nemzetközi azonosító számozása (ISBN)

MSZ ISO 3297:2000 Információ és dokumentáció. Időszaki kiadványok nemzetközi azonosító száma (ISSN)

MSZ ISO 10957:1997 Információ és dokumentáció. Kották nemzetközi szabványos azonosító száma (ISMN)

Magyarországon még nem honosított nemzetközi (ISO) szabványok és közvetlenül jóváhagyás előtt álló tervezeteik

ISRC ISO 3901:2001 Information and documentation – International Standard Recording Code (ISRC) – Hangfelvételek nemzetközi szabványos kódja

ISRN ISO 10444:1994 Information and documentation – International Standard Technical Report Number (ISRN) – Műszaki (kutatási) jelentések nemzetközi szabványos száma

ISWC ISO 15707 Information and documentation – International Standard Musical Work Code (ISWC) – Zeneművek nemzetközi szabványos kódja

ISAN ISO/FDIS (Committee Draft) 15706: 2002 – Information and Documentation – International Standard Audiovisual Number – Audiovizuális dokumentumok nemzetközi szabványos száma (A nemzetközi szabvány végleges szövegének megjelenése 2002-ben várható)

Tárgyalás alatt álló nemzetközi szabványjavaslatok

V-ISAN ISO/WD (Working Document) 20925-1 – Information and Documentation – Version Identifier for Audiovisual Works (V-ISAN) – Part 1: Format and use) 2001-11-21. Az audiovizuális dokumentumok verzióinak azonosítója 1. rész: Formátum és használat Folytatása, a V-ISAN 20925-2 az azonosító regisztrációs eljárásának szabályait fogja tartalmazni)

ISTC ISO/WD 2104 Information and Documentation – International Standard Textual Work Code (ISTC) Szöveges művek nemzetközi szabványos kódja (elsősorban szerzői jogi célokra készül és tartalmazza a megfelelő dokumentumfajta nemzetközi szabványos azonosító számára [kódjára] való hivatkozást is.)

Amerikai szabvány

SICI ANSI/NISO Z39.56-1996 Serial Item and Contribution Identifier – Időszaki kiadványok részegységeinek és közleményeinek (cikkeinek) azonosítója. (Hasonló azonosító szabványtervezetét dolgoztak ki [BICI – Book Item and Contribution Identifier] a könyvek részeinek: cikkeknek, tanulmányoknak ISBN-re alapozott azonosítására. Ez jelenleg tervezet-szakaszban van).

Az azonosító számok és kódok családfájának másik ága **az interneten elérhető (hálózati) elektronikus dokumentumokat, illetve azok elérésének útját azonosítja**. Ilyen mindenekelőtt az URL (Uniform Resource Locator, Egységes helymeghatározó), amit minden internethasználó jól ismer. Az URL a weblap (általánosítva az erőforrás) **teljes címe az interneten**.

- Legelső eleme specifikálja azt a protokollt (hálózati szabványt), melyet használni kell az eléréshez. A weben alkalmazott hipertext adatátviteli szabvány esetében ez a http (Hypertext Transfer Protocol).
- Ezt követi a “www”, mely azt adja meg, hogy a web tartománynevről van szó.
- Ami utána következik, az a tartomány vagy doménnév (domainnév). Ennek hierarchikusan legmagasabb szintű eleme a hierarchia jellegét (például “.com” a kereskedelmi, “.edu” az oktatási, “.gov” a kormányzati) vagy a földrajzi helyet (például “.hu” vagyis Magyarország) adja meg, előtte pedig az ezen belüli tartomány(ok) neve szerepel. Például a <http://www.oszk.hu/ujdonsag/tezaruruj.html> címben a teljes doménnév az “oszk.hu” rész.
- Utána következhet / jelekkel elválasztva az útvonal, melyen a tartományon belüli weblapok elérhetők (ez az előbbi példában az “ujdonsag/tezaruruj.html” rész).

Mint ismeretes, az URL mint azonosító “betegsége”, hogy érvényessége a hálózati dokumentumok kiszámíthatatlan életútja és a doménnevek változása miatt szerfölött bizonytalan. Tervezetek készültek az URL-ek változásainak nyomon követésére irányuló programokra, illetve kódokra (PURL – Permanent Uniform Resource Locator, illetve URN – Uniform Resource Name).

Jó, ha tudja a kereső, hogy az URL hierarchikus felépítése következtében **az ETO-jelzethez hasonlóan használható fel keresésre**, mert az útvonal lépésenkénti törlésével egyre magasabb hierarchiaszintű weblapok érhetők el a tartományon belül. Ha tehát nem találjuk a keresett weblapot

az addig ismert címen, az URL fokozatos jobbról végzett csonkolásával eljuthatunk ahhoz a honlaphoz, melynek alárendelt weblapja a keresett weblap volt. Arra kell ügyelni, hogy az URL-t mindig a / jeleknél rövidítsük. Ezáltal fokozatosan kaphatjuk meg a “gazdadokumentumok” URL-jeit, melyekben könnyen akadhatunk a keresett dokumentumhoz hasonló dokumentumokra irányuló elágazásokra. Persze azt is tudni kell, hogy az így elért honlap nincs feltétlenül tartalmi kapcsolatban azokkal az internet-dokumentumokkal, melyeket ugyanennek az URL-nek hosszabb változata azonosít.

Az internetforrások “címe”, az URL mint azonosító kód általában csak egészében használható fő keresésre, noha várható, hogy előbb-utóbb megjelennek olyan keresőeszközök, melyekkel a részleteik alapján is kereshető. Például sejthető, hogy ingatlanközvetítő cégek URL-jei az “ingatlan” szórészt tartalmazzák, ennek alapján elég gyorsan lehetne kiválogatni az interneten a dokumentumaikat. Erre a célra már vannak eljárások: számos keresőszolgáltatásban lehetséges a már megtalált rekordok URL-jeihez hasonló URL keresése.

Az azonosító számokon kívüli, szabványosított elemek szerinti információkeresésről szólva először megemlítjük, hogy nemzetközi (és magyar) szabványok határozzák meg az **országnevek, illetve a nyelvek kódjait** is, nevezetesen az

MSZ EN ISO 3166–1:2000 Országok és igazgatási egységek nevének kódjai. 1. rész: Országnevek kódjai (ISO 3166–1),

illetve az

MSZ 3400:2001 A nyelvek nevének háromjegyű betűkódjai,

amelyek közül az országnevek kódjainak szabványa a megfelelő nemzetközi szabvány honosított változata, míg a nyelvkód szabványa a vonatkozó nemzetközi szabvány felhasználásával készül, de az ott felsorolt nyelvek és kódok közül csak a hazai könyvtári gyakorlatban vélhetően előforduló kódokat tartalmazza.

Az eddigiekből is nyilvánvaló, hogy az azonosító számoknak az esetek túlnyomó többségében akkor van jelentősége, ha meghatározott dokumentumot keresünk. Kivételek bizonyos szempontból az **ISBN-hez hasonló “hierarchikus” felépítésű azonosító számok**, amelyeknek struktúrájából (lásd alább) következik, hogy – például az ISBN esetében – nemcsak valamely meghatározott ISBN-nel azonosított könyvet (pontosabban annak adott kivitelű és árú változatát [változatait] lehet ISBN-nel keresni, hanem egy ország (nyelvterület) illetve egy kiadó minden kiadványát is. Az **ISBN szerkezete** ugyanis az, hogy a mindenkor tízjegyű szám négy részből áll:

- az ISBN nemzetközi központja által az ország vagy nyelvterület megjelölésére, annak, illetve könyvkiadásának méretével fordított arányban meghatározott számú számjegyből,
- az illető ország vagy nyelvterület kiadói számára az ország ISBN ügynökségének (Magyarországon az Országos Széchényi Könyvtár) hatáskörében – a kiadó által a tapasztalati adatok alapján kiadott könyvek számával fordított arányban – meghatározott számú számjegyből²³,
- az illető kiadó által kiadott könyvek részére, általában az ország ISBN ügynökségének feladatait ellátó intézmény által, a kiadó bejelentése alapján meghatározott értékű, a fentiek után a 9 számjegyből megmaradó számú számjegyből,
- az – e célra meghatározott algoritmus alapján számított – egyszámjegyű ellenőrző számból.

Ebből következik, hogy az ország (nyelvterület) számából és a kiadó számából álló, jobbról csonkolt ISBN-nel **kereshetők egy kiadó kiadványai**. Elvben nem kizárható – természetesen más ismérvvel “és” kapcsolatban – egy ország (nyelvterület) területén kiadott kiadványok keresése sem, de erre célszerűbb a későbbiekben még említendő országnév kód alkalmazása. A kiadók (országoként/nyelvterületenként meghatározott) azonosítói megállapíthatók a “Publishers

²³ Ez azt jelenti, hogy a sok könyvet kiadó kiadók rövidebb (kevesebb számjegyből álló kiadóazonosítót kapnak, és ezért több könyvre van számuk, mint a kisebbeknek (akárcsak az országoknál). Egy kiadónak újabb azonosítója is lehet, ha a korábbi már “betelt”.

International ISBN Directory” c. kiadványból.²⁴ A kották nemzetközi szabványos azonosító száma, az ISMN az azonosító szám felépítése tekintetében az ISBN-hez hasonló, de csak kilenc számjegyű és az ország (nyelvterület) megjelölésére szolgáló számjegyeket nem tartalmaz.

A legfontosabb alkalmazás azonban természetesen – mint minden egyéb azonosító számnál – az ismert **ISBN-nel ellátott könyv keresése**. Ez gyakorlatilag minden számítógépes könyvtári rendszerben lehetséges. Itt említendő meg, hogy ha a katalogizáló könyvtáros tudja, hogy az ISBN hibás, a katalogizálási szabályok szerint mind a hibás, mind a javított (helyes) ISBN-t is le kell írnia. Ezt (a megfelelő almezők tartalmát célszerűen “vagy” kapcsolattal összekapcsolva) a keresésnél indokolt figyelembe venni.

Az ISBN szerinti keresésnél gyakran eredményesebb az illető ország ISBN központjához, “ügynökségéhez” fordulni.

A másik típusú azonosító szám példája az ISSN. Az adott című, pontosabban az adott kulcscímű – vagyis adott főcímű, illetve a generikus című időszaki kiadványok számára (adott szabályok szerint) megállapított című – időszaki kiadvány részére az illető ország számára meghatározott számtartományból, az ország nemzeti ISSN központja (ügynöksége; Magyarországon az Országos Széchényi Könyvtár) megállapított ISSN semmi másról nem mond semmit, csak az a jelentősége, hogy az adott kulcscímű időszaki kiadványt azonosítja. Az ISSN nyolc, kötőjellel két négyes számcsoportra osztott számjegyből áll, amelyek közül a nyolcadik az ellenőrző szám. Mivel “nem beszélő” jellegű, nem használható az információkeresés folyamatában másra, mint az adott időszaki kiadvány keresésére. Az ISSN International Centre (Nemzetközi ISSN Központ) által gondozott időszaki kiadvány-világ Bibliográfia online, illetve CD-ROM-on is rendelkezésre áll, és mind az adott ISSN-hez tartozó időszaki kiadvány adatainak, mind az adott kulcscímű időszaki kiadványt azonosító ISSN-nek a keresésére alkalmas.²⁵

A további azonosító számok szerinti keresést nem ismertetjük, mert az nem tér el a fentiekben leírtaktól. A könyvtári gyakorlatban lényeges az adott dokumentumfajta nemzetközi azonosító számának vagy akár a kiadó, illetve közreadó által adott **azonosító számnak a rögzítése a megfelelő mezőkben**, és e lehetőségek figyelembe vétele a kereséskor, illetve a kereső figyelmének felhívása e lehetőségére a súgók útján.

A nemzetközi és hazai szabványok szerinti kódok: **országnevek kódjai** – a megjelenés helyének kódja és a **nyelvkódok** – a dokumentum nyelvének kódja) a keresés lényeges eszközei, ha általában nem is önmagukban, hanem más ismérvekkel kombinálva. Alkalmazásuk – különösen a nyelvkódoké – a keresés többi ismérveivel “és”, illetve “és–nem” kapcsolatban gyakori és hasznos, mindenkor, amikor adott nyelvű, illetve adott országban megjelent (tehát nem szükségképpen az adott országgal foglalkozó) dokumentumot keresünk. Természetesen e kereshetőséghez az szükséges, hogy ezeket a kódokat (amelyek az adatformátumok – mind a HUNMARC, mind a USMARC 008 hívójelű mezőjének meghatározott pozíciójában jelennek meg) az adott könyvtári rendszer kereshetővé tegye (ami ezeknek az ismérveknek az esetében általában megtörténik).

Léteznek egyéb, az **adatformátumok említett 008 mezőjében** szintén helyet kapott **kódértékek**, amelyek főbb dokumentumfajtánként igen sok jellemzőnek a kódját (meghatározott pozíción elhelyezett értékét) adják meg, igen speciális – ha nem is mindig teljesen logikus – rendszerben. Ha a (könyvtári) információkereső rendszer erre lehetőséget ad (sajnos ritkán, de terjedő tendenciával), akkor e jellemzők szerint is keresni lehet, természetesen az esetek túlnyomó többségében nem kizárólagos keresési szempontként, hanem “és” kapcsolatban. Ilyen szempontok (ismérvek) – egyebek között – a következők:

- műfaj (dráma, esszé, levél stb.);
- bibliográfia, szótár;

²⁴ Publishers’ Interantional ISBN Directory / International ISBN Agency. – 16. ed. – München: Saur, 1989– CD-ROM változata: Publishers’ Interantional ISBN Directory Plus (CD-ROM)n / International ISBN Agency. – 1996/97– New Providence, N. J.: Bowker, München: Saur, 1996–.

²⁵ ISSN Compact (CD-ROM) / International Serials Data System. – 1992– Paris: Centre International Denregistrement des Publications en Série: Chadwick– Healey France, 1992– ISSN 1018–4783

- életrajzi kiadvány;
- kormányzati kiadvány;
- rendezvény kiadványa;
- alkalmi kiadvány;
- ún. intellektuális szint (oktatási dokumentum, ifjúsági irodalom, ismeretterjesztő irodalom stb.);
- adathordozó (mikrofilm, vakírás, bibliofil kiadvány, különlenyomat stb.);
- illusztráltság (térkép, arckép, tervrajz, címer stb.).

Ezekon kívül még számos – dokumentumfajtként részletezett, az adott dokumentumfajta jellemző – ismérv is kereshetővé tehető a 008 mező tartalmából. Ezeket az adatelemeket azonban nem mindig és nem következetesen adják meg (különösen akkor, ha a leírás olyan könyvtári rendszerben készül, amelyik nem az adatformátumra épül), ezért ilyenkor a kódértékek szerint végzett keresés lényegesen kevesebb találatot eredményez, mint ahány valójában lehetne (lásd a keresési példát a 11. ábrán).

Itt említendő még meg, hogy a 008 mező minden esetben megadott, és keresés szempontjából is igen fontos, ezért a mutatók összeállításakor – helyes eljárás esetén – mindig figyelembe vett adata a megjelenési év (évszám) adata a 7–10 pozíciókon.

6.3 Dokumentumok keresése egységesített besorolási adatelemek szerint

6.3.1 Besorolási adatok

Besorolási adat a dokumentumokról készülő bibliográfiai tételek bármely adata lehet, amely **alkalmas a bibliográfiai tétel rendezésére, illetve keresésére**. Így például besorolási adat lehet a dokumentum főcíme, megjelenési éve vagy bármely olyan adata, amely szerinti rendezésnek, illetve információkeresésnek értelme van. Nem tekintjük besorolási adatoknak a bibliográfiai tétel azon adatait, amelyeknek ugyan lényeges tájékoztatósi funkciója lehet, de amelyek alapján a tétel és a benne leírt dokumentum keresésének nincs értelme. Ilyenek például a fizikai adatok. Mindazonáltal mindazoknak az adatoknak, amelyeknek közlésmódja valamilyen mértékben szabványosított (még ha az azonosíthatóság oltárán a bibliográfiai szabványosítás időnként fel is áldozza a rendezési/keresési szempontokat), adott esetben más (többnyire egységesített) besorolási adatokkal együtt fontos szerepük lehet a keresésben.

Ugyanakkor a besorolási adatok szerinti rendezés vagy keresés (a továbbiakban csak a keresést említjük) gyakran nem biztosítja a keresés egyértelműségét. A bibliográfiai adatok jelentős része – még ha szabványosítottan írjuk is le őket, erre nem alkalmas. Példának talán elegendő arra hivatkozni, hogy igen sok azonos személynevű ember él, tehát a szerzők neve szerinti keresés még korántsem teremt biztonságot abban a tekintetben, hogy az általunk keresett szerző művét/műveit kapjuk eredményként. Hasonlóképpen, hiába lényeges besorolási adat a dokumentumok címe (főcíme), a főcím szerinti keresés a főcím különféle (köztük nyelvi) változatai miatt nem mindig ad teljes eredményt, vagy éppenséggel eredménytelenül, esetleg téves eredménnyel végződik.

Mindez arra vezetett, hogy a gyakorlati igények kiszolgálása érdekében a besorolási adatok számottevő részét **egységesítették**. Az **egységesített besorolási adat** rendeltetése szerint azt szolgálja, hogy valamely személy műveit, valamely testület által közreadott műveket, valamely mű kiadásait, valamely személlyel, testülettel, földrajzi területtel vagy művel kapcsolatos dokumentumokat e személyek, testületek nevének, e művek címének vagy földrajzi területek megnevezésének – egységesített alakja szerint lehessen keresni. Az egységesített besorolási adatok megjelenítésének is szigorú szabályai vannak az egyértelműség érdekében; például e szabályok meghatározzák az egységesített besorolási adatok részeit és ezek közlésmódját. A számítógépes megoldások esetében ezeknek az előírásoknak betartását – bár azok ettől függetlenül legalábbis az

adatok megjelenítésekor kötelezők vagy legalábbis ajánlottak – adott esetben helyettesíti az adatok adatsere-formátumban történő rögzítése és e szerinti kereshetővé tétele.

Az egységesített besorolási adatok nyilvántartása és használata és az adattárak egységesítése érdekében az egységesített besorolási adatokból és ugyanazon besorolási adat egyéb alakjaiból, valamint – a besorolási adat jellegétől függő – kiegészítő adatokból úgynevezett **besorolási adattárakat** (“authority” fájlokat) hoznak létre. Ezek tartalmazzák a besorolási adat egységesített alakját, az annak egyértelműségét biztosító kiegészítő adatokat (például személyek születési és halálozási évszámát, testületek székhelyét), ugyanazon besorolási adat egyéb alakjait (a hagyományos katalógusban “lásd utalók”), egyes, többé-kevésbé kivételes esetekben, mint például az egységesített besorolási adatnak az idők folyamán végbement változásai esetén a lásd még, lásd előbb vagy lásd utóbb értelmű utalókat és egyéb tájékoztató, illetve technikai adatokat.

A számítógépes rendszerekben a besorolási adattárak egy-egy besorolási adatról készülő rekordjai kapcsolódnak mindazon dokumentumok bibliográfiai tételeihez, amely tételekben az adott besorolási adat előfordul. Ez azt teszi lehetővé, hogy a besorolási adat változása, és – következésképpen – a besorolási rekord változása esetén a megfelelő adat program szerint automatikusan megváltozik mindazokban a bibliográfiai tételekben, amelyek az illető besorolási adatot tartalmazták. Ezt az eredeti – angol – terminusok szerint (global change vagy global update) **általános módosításnak** vagy **általános korszerűsítésnek** nevezhetjük.

A bibliográfiai tétel besorolási adatai megválasztásának és közlésük módjának szabályait az MSZ 3440/1–5 sz. 1979-ben hatályba lépett szabványok (szabványcsalád tagjai) tartalmazzák. A 3440/1 sz. szabvány a szabványcsaládban használt fogalmak meghatározásait adja; a többi szabvány számát és címét a megfelelő besorolási adattal foglalkozó részben adjuk meg. Teljes bibliográfiai hivatkozásukat itt nem közöljük (hozzáférhetőek a Könyvtári Intézet Könyvtártudományi Szakkönyvtárában). E szabványok rendelkezései ma is érvényesek, bár egyes (például központosási) szabályok a számítógépes katalógusok készítése, a számítógépes besorolási adattárak és az adatsere-formátumok alkalmazásának körülményei között jelentőségüket, sőt értelmüket veszítették, illetve főként csak a besorolási adatok megjelenítésekor bírnak – az adott rendszerben hozott döntéstől függően – jelentőséggel.

A vonatkozó szabványok megkülönböztetik az egységesített besorolási adatok rendszóelemét, egyéb(név)elemeit és kiegészítő adatait. Az azonos besorolási adatokat megkülönböztethetőségük és kereshetőségük érdekében az adathoz járuló kiegészítő adatokkal kell ellátni. A rendszó és az egyéb névelem(ek) megkülönböztetésének a számítógépes keresés szempontjából nincs jelentősége; a szabványok e megkülönböztetést a cédulakatalógusok és jegyzékek rendezési előírásai miatt alkalmazzák. A kiegészítő adatokra esetenként visszatérünk.

A nem egységesített besorolási adatokat a megfelelő egységesített besorolási adattal (például egy személy egységesített nevét a névnek a szerző által használt és/vagy egyéb változataival) a besorolási adattárak útján gépi úton is nyomon követhető utalók kapcsolják össze. Ugyanez vonatkozik – szükség szerint, értelemszerűen – az egységesített besorolási adatok rendszón kívüli elemeire (például az elsődleges besorolási adaton kívüli egyéb névalakokra). Mind az egységesített besorolási adatok kezelése, mind az utalók tekintetében a számítógépek alkalmazása új, egyszerűbb és biztosabb megoldásokat tett és tesz lehetővé, anélkül azonban, hogy alapjában befolyásolják az egységesítéssel kapcsolatos követelményeket.

Az alábbiakban röviden áttekintjük az egységesített besorolási adatok körét és az azokkal való keresés lehetőségeit.

6.3.2 Személyek nevei

Napjainkban már kétségtelenül elsődlegessé vált a dokumentumok tárgykör szerinti keresése. A dokumentumok és szerzőik számának növekedése, az ismeretek robbanásszerű bővülése, valamint a kommunikáció méreteinek és összetettségének fejlődése a múltnak adta át azt az időt, amikor valamely tárgykör kereséséhez elég volt a téma legnagyobbjainak nevét ismerni. A szerzők (és közreműködők, vagyis a személyek) az információkeresés szempontjából mégis – ma is – többet

jelentenek, mint annak lehetőségét, hogy személyi bibliográfiákat lehessen összeállítani. A korábbiakban írottakból is kiderül, hogy gyakran éppen adott szerző vagy adott szerző adott műve a kulcs a tárgykör további eredményes kereséséhez. Végül pedig – az eredményes tárgyköri keresés után – a talált műveket meg is kell keresni a könyvtárban, amihez a vonatkozó személynevek – és azok egységesített közlése elengedhetetlenül szükségesek.

Mindenekelőtt lényeges, hogy a személyek nevét (egyébiránt a testületekét is) a bibliográfiai tétel kétszer tartalmazza. Maga a bibliográfiai leírás – amelynek fő funkciója az adott dokumentum (tehát nem vagy nem elsősorban a mű, hanem az adott fizikai egység) azonosítása – ennek az azonosításnak érdekében a személyek (szerzők, közreműködők) nevét az úgynevezett szerzőségi közlés (szerzőségi adatok) keretében abban a formában és mindennemű kiegészítés nélkül közli, ahogy azok a leírás tárgyát képező dokumentumon megtalálhatók. Mivel ez számos okból nem lehet egységesített alak (nem beszélve arról, hogy több azonos nevű szerző létezik/létezhet), az adat nem alkalmas adott személy műveinek vagy adott személy közreműködésével készült műveknek a személy neve szerinti keresésére.²⁶ A bibliográfiai leírásban közölt szerzőségi közlés még csak nem is besorolási adat, ami – korszerű rendszerek alkalmazásaira lefordítva – azt jelenti, hogy a szerzőségi közlésben közölt személynevek szerint nem vagy – ritka kivételként – nem a teljes név hanem a név szavai szerint lehet keresni.

Az egységesített személynév használatának természetesen alapelve az, hogy **ugyanannak a személynek a nevét** ugyanabban a rendszerben (persze kívánatosan minden rendszerben) **mindig ugyanazon a néven kell megadni** (leírni), még hozzá abban a névalakban, amelyen a személy működését kifejtette, illetve amelyen a nemzeti hovatartozásának megfelelő bibliográfiai gyakorlat számon tartja. A kiegészítő adatok közül a legfontosabb – és az esetek túlnyomó többségében a személyt egyértelműen azonosító – adat a személy születési, valamint – ha elhunyt – halálozási évszáma. Mindezt célszerű a lehető leghitelesebb forrás alapján pontosan leírni. A pontatlanság olyan abszurd adatokra vezet mint például Móricz Zsigmond különféle születési/halálozási adatai a 9. ábrán. További, a személy egyértelmű megkülönböztetését célzó kiegészítő adatok is vannak; így a személy idősebb vagy ifjabb voltának jelzése, foglalkozása, a különböző világi és egyházi méltóságok megkülönböztetésére szolgáló adatok stb.

A személyek nevének mint egységesített besorolási (keresési) adatnak a közlésmódjáról az MSZ 3440/2–79 szabvány részletesen intézkedik. Meghatározza a személyek nevének elemeit (rendező, egyéb névelem, kiegészítő adatok); az összetett személynevek kezelésének szabályait; külön tárgyalja fontosabb nyelvenként a modern névhasználat, majd a történelmi névhasználat, az uralkodók és uralkodóházak tagjainak névhasználatát, a vallási nevek, az írói nevek, művésznevek és álnevek, a pótnevek (névhelyettesítők) besorolási adatként való használatának szabályait, valamint az azonos nevek megkülönböztetésére szolgáló kiegészítő adatokat.

Mind a személyek neveire, mind a más típusú egységesített besorolási adatoknak a számítógépes rendszerekben való használatára egyaránt vonatkozik, hogy – felhasználva az adatcsereformátumok adta megkülönböztetési lehetőségeket, nevezetesen a besorolási adatok különböző elemeinek meghatározását és sajátosságait – a gépi (online) mutatók szerkezetének és tartalmának meghatározásakor figyelembe kell venni ezeket a szabályokat.

Nem kevésbé lényeges az utalókra vonatkozó szabályok figyelembevétele, még akkor is, ha ezek mintegy automatizált nyomon követésének feladatát az egységesített besorolási adatokat tartalmazó besorolási adattárak (“authority” fájlok), illetve az azokban végzett keresések veszik át.

Bár a magyar szabvány is megemlékezik – ha összevontan is – az egyes nyelvek, illetve nemzetek névhasználatáról, megemlítendő, hogy e tárgyban további információkkal szolgál a Könyvtári

²⁶ Hasonlóképpen nem alkalmas a személyek nevének közlésére, amikor a személyről szóló mű esetében a személy nevét úgynevezett tárgyi kivetítésre alkalmas, természetesen egységesített formában kell közölni, hagyományos rendszerben melléktétel besorolási adatként, gépi rendszerben mint kereshető, a dokumentum tárgyat jelző adatot.

Egyesületek és Intézmények Nemzetközi Szövetségének keretében készült kézikönyv, a Names of persons, amelyet magyar változatban is kiadtak.²⁷

6.3.3 Testületek nevei

A testületek (ideértve a rendezvényeket is) növekvő szerepet játszanak, mint az információk létrehozói vagy feldolgozói és közreadói. Nevük – a testület jellegétől, a nemzeti (és nemzetközi) szokásoktól függően – rendkívül sokrétű szerkezetű. Ebből adódóan a hagyományos katalógusokban igen nagy súlyt kellett helyezni olyan névalakok használatára, amelyek szerint a testület (illetve a vele kapcsolatos, többnyire általa közreadott dokumentum) kereshető. A számítógépes rendszerek javították a keresés eredményességének esélyeit; ennek ellenére hasznos tudni a legfontosabb szabályokat az eredményes keresés és a kapott találatok használata érdekében. A szabvány **testületi névnek tekinti a rendezvények** nevét is. A testületi nevek közlésének alapelve ugyanaz mint a személyek esetén: ugyanazt a testületet mindig ugyanazon a néven, általában teljes, hivatalos, eredeti nyelvű nevével kell megnevezni, a névváltozatokat, így a közkeletű nevet, a nyelvi változatokat utalókban – számítógépes besorolási adattárak esetén – a testület besorolási adatrekordjában az utaló szerepét betöltő, kereshető névváltozatként kell figyelembe venni. Mivel a testületeknek (elsősorban, de nem kizárólagosan a nemzetközi testületeknek) több nyelven is van neve, továbbá a legtöbb testület teljes hivatalos neve mellett rövid névalakot is használ, ez esetben is **utalókkal**, illetve a besorolási adattárban kell lehetőséget adni ezek összekapcsolására a kitüntetett (általában a teljes, hivatalos) névalakkal. A szabvány (MSZ 3440/3) rendelkezik arról, hogy a teljes hivatalos névből milyen elemeket kell elhagyni (névkezdő névelő, a testület kitüntetései megnevezései, a tulajdonjogi formát jelölő kifejezések). A testület **székhelye a testületi név kiegészítő adata**, amit azonos nevű, de különféle székhelyű testületi nevek esetében figyelembe kell venni. A székhely adatát nem közlik **egyházak és egyházi rendek, valamint állami és helyi igazgatási szervek** esetében. Ez utóbbiak szabványos neve a működési terület nevével kezdődik és ezt követi a testület jellegét, jogállását, funkcióját jelző hivatalos név (például: Magyarország. Országgyűlés). Alárendelt testületek esetében a testületi név általában csak az alárendelt testület neve (bár ez alól is vannak kivételek).

A szabvány rendelkezik még a testületi nevek kiegészítő adatairól, amelyek a testület nevét egyértelművé teszik.

A fenti vázlatos (nem teljes) áttekintésből látható a szabályok meglehetősen, a járatlan információkeresőt gyakran zavarba ejtő sokfélesége. Ebből a sokféleségből következik, hogy testületek nevének keresése esetén a legcélszerűbb a testület nevének szavai szerint “és” kapcsolattal keresni. Ebben az esetben – ha a kereső például nincs tudatában az igazgatási szervekre vonatkozó szabályokkal, de például az “Oktatási Minisztérium” szavakkal keres, akkor bizonyára megtalálja a keresett tételeket. Ha pedig a szabványos névalakot kívánja megtalálni, azt a találatból láthatja, illetve szükség esetén a besorolási adattárból megtudhatja a testületnek a szabvány szerinti nevét és annak kiegészítő adatait.

Amint az a fentiekből, de különösen a szabvány szövegéből látszik, a szabályok számottevő része a manuálisan szerkesztett katalógusok és az azokban való keresés céljait szolgálják. Mindazonáltal a nevek kezelése nem zavarja a számítógépes keresést.

A magyar szabvány főbb rendelkezései a vonatkozó nemzetközi tanulmányon alapulnak.²⁸

²⁷ A személyek neve mint a katalógusok besorolási adata az egyes nemzetek gyakorlatában, ford. Orbán Éva; a magyar kiadás bevezetését írta Fügedi Péterné; szerkesztette Varga Ildikó; közreadja az Országos Széchényi Könyvtár, Könyvtártudományi és Módszertani Központ. – Bp. OSZK–KMK, 1986. – 219 p.

6.3.4 Címek

A címekről mint besorolási adatokról az MSZ 3440/4–86 szabvány rendelkezik, amelyik az MSZ 3441 szabványcsalád hatályát vesztett, azonos számú, de 1981-ben jóváhagyott változatának helyébe lépett. A címek többségének esetében a besorolási adatként használt cím megegyezik a dokumentum főcímével. A besorolási adatként használatos cím elsősorban akkor tér el a leírt főcímtől, ha a mű eredeti címe eltér a leírt dokumentum címétől. **Besorolási adatként ugyanis mindenkor az egységesített, eredeti címet** kell leírni. Manuális megoldás esetén természetesen utalót kell készíteni a leírt dokumentumon közölt címről. Ha – számítógépes megoldás esetén – a címekről is készül besorolási adattár, akkor az eredeti címről készül a besorolási adattár megfelelő rekordja és abban címváltozatként (utalóként) szerepel a dokumentumnak a leírásban megjelenő címe.

A fenti általános szabály azt is jelenti, hogy a fordításban megjelent művek besorolási adatként közölt eredeti címe a mű **eredeti nyelvű** címe.

Besorolási adatként kell közölni egyes esetekben a dokumentum (mű) **átültetett címét**, amely a műnek a bibliográfia (esetünkben az adatbázis) nyelvének megfelelő bibliográfiai gyakorlatban meghonosodott cím. Ezt az eljárást kell követni a vallások szent könyvei, az ismeretlen szerzőjű (anonim) ókori, középkori és humanista művek és – eltekintve azoktól az esetektől, amikor a zenemű címe műfaji megnevezés vagy a zeneszerző azonos műfajban írt művei között sorszámmal vannak ellátva a zeneművek esetében. Ez (lásd alább) a megállapodásos cím.

Megállapodásos cím az egységesített cím akkor, ha valamely gyűjteménynek nincs megkülönböztető címe. Ilyenkor “Művek” szó vagy a műfajt megjelölő más szó a megállapodásos cím.

Információkeresési szempontból az egységesített címek gyakran segítik a keresett dokumentum megtalálását, de célszerű cím szerinti keresésnél (sőt már az indexek létrehozásánál) a bibliográfiai leírás szerinti főcím (és esetleg az egyéb cím, valamint az egységesített cím) együttes figyelembe vétele. Lefordított művek esetében mindenkor célszerűbb először ez eredeti (vagyis egységesített) cím keresése, bár e tekintetben a szerző és a főcím vagy annak szavai alapján megtalált bibliográfiai tétel is eligazítást adhat.

6.3.5 Földrajzi nevek

Természetesen viszonylag ritkán – bár korántsem kizárhatóan – fordul elő, hogy valakit minden érdekel, ami egy meghatározott területre vonatkozik. Ennél sokkal gyakoribb, hogy valakinek a földrajzi területre mint kereső szempontra valamilyen tárgykörrel kapcsolatban van szüksége (például balatoni halászat). Természetesen ilyenkor kereshet a földrajzi terület deszkriptora vagy osztályozó jelzete szerint is. Földrajzi nevek szerinti kereséskor az országnévkódok két okból sem használhatóak: egyrészt e kódok általában a dokumentum származási helyét jelölik, és nem azt a helyet, amiről a dokumentum szól, másodsorban mert a kódok (legalábbis a bibliográfiai gyakorlatban használt kódok) nem terjednek ki az országok részeire (illetve ha igen, akkor csak igazgatási egységekre, mint amilyen a megye, grófság, régió), végül pedig nem terjednek ki természetföldrajzi egységekre, mint például vizekre (Balaton, Csendes-óceán), hegységekre (Mátra, Kilimandzsáró) stb. Ezért a földrajzi nevek szerinti kereséshez szükség volt a földrajzi név mint besorolási adat bibliográfiai célokra történő egységesítésére. Mivel a földrajzi név **külön egységesített besorolási adatként nem létezik**, az egységesített földrajzi név használatára akkor

²⁸ Corporate headings; their use in library catalogues and national bibliographies. A comparative and critical study / by Eva Verona. – London: IFLA Committee on Cataloguing, 1975. – XIV, 24. p.

van szükség, ha a személynév vagy testületi név egységesített besorolási adatának részeként, **kiegészítő adatként** a szabványok szerint földrajzi nevet kell megadni. Emellett az egységesített földrajzi névre vonatkozó szabályokat veszik alapul – vagy kellene alapul venni – **tezauruszok és tárgyszórendszerek megfelelő adatainak megválasztásakor** is. Ez természetesen nem zárja ki a földrajzi név egységesített alakja szerinti keresést, de annak lehetőségeit korlátozza.

A vonatkozó szabvány a besorolási adatok szabványcsaládjának **MSZ 3440/5–79 számú tagja**. Rendelkezéseinek szerkezete hasonló az egységesített besorolási adatok képzését és használatát szabályozó többi szabványhoz. Alapelve, hogy az adott földrajzi terület megnevezésére mindig ugyanazt a nevet kell használni.

Az általános elv itt sem tér el a többi egységesített névre vonatkozó szabályoktól: a földrajzi helyet mindig **ugyanazon a néven (az egységesített néven)** kell megnevezni. A szabvány rendelkezése szerint az egységesített földrajzi név a térképészeti gyakorlatban elfogadott magyar nyelvű vagy részben magyar nyelvű vagy eredeti hivatalos nyelvű alakja. Ennek közlésmódjára a szabvány részletes előírásokat tartalmaz.

Az **országok és hasonló területek** földrajzi nevét az ország közkeletű nevének magyar nyelvű alakjaként kell megadni. A teljes, hivatalos államnév és az eredeti nyelvű név – számítógépes megvalósítás esetén akkor, ha a rendszer ilyent tartalmaz – a földrajzi nevek besorolási adattárában kaphat helyet névváltozatként (“utalóként”). Ha nincs magyar országnév, az eredeti nyelvű hivatalos országnevet vagy többnyelvű országok esetén ezek egyikét kell használni. Az országnevekkel kapcsolatban megjegyzendő, hogy az országnévkódok korábban említett szabványában megadott országneveket kritikával kell felhasználni, mert azok az eredeti nemzetközi szabványhoz ragaszkodva készültek és nem mindig helytállóak.

Az **egyéb igazgatási területek** (megye, grófság, szövetségi államnév) egységesített neve az egységesített név eredeti nyelvű hivatalos nevüknek az igazgatási terület típusát jelölő szavak és az esetleges ehhez tartozó viszonyzó nélküli része, ha azonban a típus az igazgatási terület nevének elengedhetetlen része, akkor nem hagyható el. **Helységnévként** a helységek nevének eredeti, hivatalos nyelvű alakját (vagy több ilyen alak esetén ezek egyikét) kell megadni. Helységek részeinek egységesített neve a helység egységesített nevéből és a városrész megkülönböztető megjelöléséből vagy nevéből áll.

Megváltozott földrajzi nevek korábbi alakját utalók (gépi rendszerekben a besorolási adattárban) kapcsolják össze az érvényes névvel. Több azonos nevű földrajzi egység esetében a kiegészítő adat a megfelelő ország neve, vagy kódja, vagy az igazgatási terület teljes, vagy rövidített alakja adható meg.

Gazdaság-, illetve természetföldrajzi nevek egységesített alakja a magyar nyelvű név (az eredeti névről készült utalóval). Ha nincs magyar név, az eredeti nevet kell használni. Szükség esetén kiegészítő adatként az ország nevét adják meg.

Azonos földrajzi nevek megkülönböztetésére kiegészítő adatként országok esetében (ahol elsősorban a történelem során változott név fordul elő – az államformát, kisebb egységek és természetföldrajzi helyek neve esetében a földrajzi hely típusát adják meg **kiegészítő adatként**).

7. Tárgyköri keresés. Az információkereső nyelv szótárainak használata

7.1 Keresés kötött szótár használata nélkül (“szabad tárgyszavas keresés”)

Mielőtt e téma lényegére térnénk, igyekszünk megmagyarázni, hogy a fejezet címének második részét miért adtuk meg, és ha már megadtuk, miért tettük zárójelbe. Az ok egyszerű: a kifejezés széles körben elterjedt (ezért tüntettük fel) de ugyanakkor helytelen (ennek előrejelzésére szolgál a zárójel). Igaz ugyan, hogy a **“keresés a szövegben”** – ami egyébként a folyamat lényegét kifejező kifejezés – az esetek jelentős részében a dokumentum tárgyára irányul, de a **“tárgyszó”** kifejezés már foglalt, mégpedig a tárgyszavas rendszerek szavaira; nem lenne tehát célszerű a szövegben előforduló, nem egységesített, nem szabványosított szavak, kifejezések megnevezésére használni.

Ha mindenképpen ilyen megnevezést kellene választani, akkor **“szabad szövegszó”** lenne alkalmas, ez viszont tautológia, mert a dokumentum szövegének szavai természetüknél fogva **“szabadok”**, hiszen nem kötöttek, nem szabályozottan, nem eleve a keresés céljaira **“készülnek”** (írják le őket). Valójában arról van szó, hogy a tárgyköri keresés számára az információkereső nyelvek – Madách megfogalmazásával élve – **“túl szűk és túl tág lombikok”**. Az információkereső nyelvek még folyamatos karbantartásuk esetén sem képesek lépést tartani az ismeretek fejlődésével; **új fogalmakat kifejező új kifejezések csak fokozatosan, “lassan” kerülnek be az információkereső nyelvek szótárába** – mindegyik fajtájuk esetén, de különösen a hierarchikus szerkezetű nyelvek esetében. Másrészt egyes kifejezések annyira speciálisak, hogy ezért nem kapnak helyet az információkereső nyelv szótárában, és nem is használják őket a dokumentum tárgyköri indexelésekor. Végül az információkereső személy – legyen ez akár a könyvtáros, akár az információs bróker, különösen pedig maga az információt igénylő szakember – **gondolkodásmódja, fogalomkészlete** nem szükségképpen felel meg az információkereső nyelv szerkezetének és tartalmának, különösképpen pedig az indexelő vagy osztályozó szakember szemléletének, vagy éppenséggel ismeret-anyagának.

Ezért már a gépesített információkereső rendszerek létrejöttének korai szakaszában voltak olyan törekvések, hogy az osztályozó jelzetek, tárgyszavak, deskriptorok szerinti keresés mellé egyéb keresési módok csatlakozzanak. Már e rendszerek alkalmazásának kezdeteikor nemcsak a dokumentumnak az információkereső nyelv szavaiból álló keresőképét **invertálták** és tették ezáltal az invertált fájlok alkalmazásával kereshetővé, hanem a **dokumentumok címének** (vagy címeinek és ezúttal korántsem csak az egységesített címeknek) önálló **értelemmel bíró szavait**. Ezt úgy tették, hogy a cím minden szavát invertálták, vagyis a dokumentum azonosítójával együtt felvették a keresést szolgáló invertált fájlba, az önálló értelemmel nem bíró és ezért **“tiltott szavak szótárába”** (**“stopszótárba”**) (példáját lásd a 2. táblázatban) felvett szavak kivételével. Később ez a gyakorlat kiterjedt a **dokumentumok kivonataira** (referátumaira) vagy annotációira, adott esetben a témát összefoglaló első bekezdésekre is. Elképzelhető, hogy – ha ezt a technika fejlődése megengedi (és miért ne engedné meg), a távolabbi jövőben invertálni fogják a vélhetően információkeresés tárgyát képező, bizonyos kritikus méretet meg nem haladó és a használók által várhatóan keresett dokumentumok minden önálló jelentéssel bíró szavát. Ami viszont a **címek szavai szerinti keresést** illeti, ezt szinte minden könyvtári rendszer már ma megvalósítja, mivel ez nem is csak a tárgyköri keresés céljait szolgálja, hanem – a szerző ismeretében és a szerző nevével **“és”** kapcsolatban keresve vagy a szerző neve nélkül – azon **konkrét dokumentum keresésének célját is**, amelynek létéről a kereső tud.

A szöveg szavai szerinti keresés jelentős **előnye**, hogy az újonnan létrejövő, az adott rendszerben alkalmazott információkereső nyelv szótárában még nem, vagy egyáltalán nem szereplő fogalmak is kereshetők. Ugyanakkor e keresés számottevő **hátránya** – bár erről hívei ritkán szólnak – hogy az adott keresés tárgyának tartalmi bővítésére vagy szűkítésére, vagy valamilyen irányba eltérésére (legalábbis információkereső nyelvi **“segédletek”** nélkül) ez a technika, a megfelelő eszköztár híján nem alkalmas.

A legnagyobb adatbázis-szolgáltatók gyakorlatában természetes **“polgárjogot”** nyert a címek és referátumok szövegében végzett információkeresés. Finomítására különféle eszközök állnak rendelkezésre. A szöveg szavai szerinti keresésnél figyelembe kell venni egyfelől, hogy a szavak különféle toldalékokkal, adott esetben előtagokkal, különféle névszói, ragozott alakokban szerepelhetnek a szövegben. Nem kevésbé jellemző, hogy a keresett szavak **“közös előfordulása”** a keresés követelménye lehet. Mindezekre a célokra különféle **csonkolási, maszkolási és közelség-meghatározási technikák alakultak ki**. Ezek jelzésére a különféle adatbázis-szolgáltatók rendszerei (akár csak a különféle könyvtári rendszerek) különféle jeleket alkalmaznak (részletesen a 3.4.2 fejezetben tárgyaljuk). Például a (jobbról) csonkolás jele a DIALOGnál **?**, az ORBITnál: **.**, a BRS-nél **\$**, amely után számjegy állhat, azt jelezve, hogy a **“levágott”** karakterek száma maximálisan mennyi lehet; például **meteor\$3** azt (is) jelenti, hogy a szó lehet meteorit, de nem lehet meteorológia.

A szavankénti szövegkeresés nem mindig ad megfelelő eredményt. Ha például két vagy három szó szerint keresünk, akkor ezek sorrendjétől és egymástól való távolságától függően különféle találatokat kaphatnánk, amelyeknek csak egy része vagy egyik sem felel meg a keresett témának. Ezért alkalmazzák a **szóláncok szerinti keresést**, amikor nem egyszerűen szavakat, hanem együttesen (egymás után) előforduló több szót keresnek a szövegben (lásd még a 3.4.3 fejezetet). Ennek a módszernek az alkalmazása általában megkívánja, hogy először a szavakkal végzett kereséssel létrehozzunk egy találati halmazt, és abban keressünk – kötött sorrenddel – a szólánc szerint. A szólánc meghatározásánál **rögzíteni kell magukat a szavakat és azok egymáshoz képest “megengedett” elhelyezkedését**. Így például, ha roncsolásmentes anyagvizsgálat tárgykörében keresünk és a DIALOG jelöléseit használjuk, a

roncsolásmentes (**w**) anyagvizsgálat

“lánc” azt jelenti, hogy a két kifejezésnek együtt (egymás mellett közvetlenül) kell előfordulnia. A w mellé egy számjegyet írva, például:

roncsolásmentes (**2w**) anyagvizsgálat

eredményként kapjuk mindazokat a dokumentumokat ahol a két szó úgy fordul elő, hogy köztük még egy szó van, például roncsolásmentes izotópos anyagvizsgálat A (**w**) beékelt jelhez (infixumhoz) hasonló funkciójú más jeleket alkalmazva a “szóláncnak” azokat az eseteit is kereshetjük, amikor a “lánc” elemei egymástól távolabb helyezkednek el. Például a DIALOG-nál az (**f**) infixum azt jelenti, hogy a megelőző és követő szavaknak megtalálhatónak kell lenniük a szövegben, de az egy mezőn belüli szavak esetében egymástól tetszőleges távolságra. A BRS az infixumok helyett szavakat, vagy rövidítésüket használja, nevezetesen az (**adj**) jelenti azt, hogy a két szónak “egymás mellett” (szomszédosan) kell elhelyezkednie, (**with**) jelölés esetén ugyanabban a bekezdésben, (**same**) jelölés esetén pedig ugyanabban a mezőben kell előfordulnia.

A **könyvtári rendszereknek a szabad szövegkeresésre használt eszköztára nem ennyire kifinomult**. Általában a keresett szavak (időnként VAGY kapcsolatokkal fűszerezett) ÉS–kapcsolatával jelölik meg az (általában a címben, illetve címekben) keresett szavakat, hacsak egyetlen szó önmagában nem olyan jellegzetes, hogy vélhetően egyedüli keresőszóként is eredményesen lenne használható.

Lényeges a szövegben végzett keresés esetében, hogy e módszer nem teszi feleslegessé, sőt **feltételezi a természetes nyelven alapuló információkereső nyelvek szótárainak használatát**, vagy akár a tárgyszavas rendszerek tárgyszavainak és utalószavainak, vagy a teauruszok deszkriptorainak és nemdeszkriptorainak, vagy a hierarchikus nyelvek természetes nyelvű adatainak használatát. Ez azt jelenti, hogy a szövegben végzett keresésre készülő személynek célszerű a keresési utasítás összeállítására előtte megkeresni az általa keresett szót a tárgyszórendszerben, vagy teauruszban, vagy a hierarchikus rendszer táblázataiba, hiszen ezáltal olyan szinonimákra, kváziszinonimákra, vagy a szóbanforgó keresőszóval más összefüggésben álló szavakra találhat, amelyek VAGY kapcsolatban alkalmazva eredményesebbé tehetik a keresést.

A fentiekben azt állítottuk, hogy csak olyan adatelemek (mezők) szavait lehet szabadon keresni, amelyek szavaiból **invertált fájl** készül. Megemlítendő azonban, hogy egyes rendszerek felkínálnak olyan keresési lehetőséget is, hogy a keresett szónak vagy szavak kombinációjának előfordulását az összes rekord illetve meghatározott rekord-csoport megfelelő mezőjében vagy mezőiben **szekvenciálisan** (szóról szóra haladva) hasonlítják össze a rekord szavaival, egyébként az említett módszerekhez hasonlóan. Ez az eljárás azonban – legalábbis ma még – igen hosszadalmas és ennek megfelelően költséges.

Összefoglalva elmondható, hogy a **keresés a szövegben (vagy szabad szavas keresés)** már ma lényeges kiegészítője, illetve megfelelő esetekben alternatívája lehet a kötött szótár szerinti keresésnek, a jövőben pedig alkalmazása várhatóan gyors fejlődésen megy át.

7.2 Keresés kötött szótár használatával

7.2.1 A keresés lépései kötött szótár használatakor

Ha az információkereső rendszerben kötött szótárt használnak, akkor minden célzott keresést elvileg megelőz egy böngészési folyamat. Nem tudhatjuk előre, hogy pontosan milyen formában szerepel annak a fogalomnak a megnevezése, mely szerint keresni szeretnénk. Ehhez meg kell keresni a kötött szótárban (tárgyszójegyzékben, teauruszban, osztályozási rendszerben) azt a kifejezést, amelyről úgy gondoljuk, hogy a keresőfogalmunkat megnevezi.

A kötött szótárakba általában az információkereső rendszer online mutatójából léphetünk át. A mutatóban vagy még inkább a szótárakban nem egy esetben még ide-oda is kell navigálni, mert könnyen kiderülhet, hogy “nem azonos nyelvet beszélünk” az információkereső rendszerrel, és nem pontosan olyan alakú az a keresőszó, amelyet ténylegesen használnunk kell, mint amelyből kiindultunk.

Miután a szótárban ráakadtunk a vélelmezett keresőszóra, célszerű tájékozódni ennek szemantikai (jelentésbeli), – osztályozási rendszerek esetében hierarchikus – környezetéről is. Ennek feltétele, hogy a kötött szótárnak legyen némi szemantikai (vagy másszóval: paradigmatis), illetve hierarchikus struktúrája, magyarul szerepeljenek benne összefüggések, kapcsolatok a szótár lexikai egységei (a tárgyszavak, deszkriptorok, osztályozási jelzetek) között.

A szócikkben (tárgyszócikkben, teauruszcikkben, ETO-jelzet környezetében) látható összefüggések a gondolkodás szempontjából asszociációkat képviselnek. Belőlük egyrészt következtethetünk arra, hogy jól választottuk-e ki keresőszavunkat? Ha ugyanis jól választottunk, akkor a szócikkben olyan kapcsolódó kifejezéseket találunk, melyek nem mondanak ellent annak a fogalmunknak, melyre a keresést megelőzően gondoltunk.

Másrészt segítséget kaphatunk a tekintetben, hogy esetleg milyen más keresőszót érdemes még bevonni a keresésbe (szinonimákat, speciálisabb vagy átfogóbb jelentésű szavakat, jelzeteket stb.)

A keresés egyes lépései kötött szótár esetében:

- megfogalmazzuk a keresőszót a saját nyelvünkön;
- kiválasztjuk a keresőrendszerben azt az online mutatót, mely a kereséshez használandó kötött szótár lexikai egységeit tartalmazza;
- megadjuk az általunk előzetesen elgondolt keresőszót;
- megjelenik a mutatónak az a része, mely a megadott szóval vagy a betűrendben hozzá legközelebb álló szóval kezdődik;
- szükség esetén fel s alá navigálunk az online mutatóban (böngészünk);
- szükség esetén a mutatószóról áttérünk ennek a szónak a szócikkére és ebben is böngészünk;
- szükség esetén visszatérünk a betűrendes mutatóba és az előző két műveletet megismételjük;
- a végleges keresőszót kiválasztjuk, és ezzel vagy közvetlenül célzott keresést hajtunk végre, vagy beépítjük a keresőképbe (ha nem csak egyetlen keresőszóval kívánunk keresni);
- meghatározzuk (ha lehet) a találatok megjelenítési formáját, sorrendjét stb., majd a találatok elemzése alapján döntünk a keresés esetleges megismétléséről.

7.2.2 Keresés tárgyszójegyzék vagy teaurusz használatával

7.2.2.1 A kapcsolatok bevonása a keresésbe

Teauruszok és tárgyszórendszerek között használatukat illetően nincs lényegi különbség. A teauruszokban elvileg az adott ország teauruszszabványában meghatározott kapcsolatokra, olykor magyarázatokra is számíthatunk a vezérszavak alatt²⁹. A tárgyszórendszerekben a kapcsolatokat leegyszerűsítve kezelik (csak “lásd” és “lásd még” kapcsolatokat alkalmaznak), de előfordulhat,

²⁹ Magyarországon az MSZ 3718-as szabvány, nemzetközi szinten az egynyelvű teauruszokra az ISO/IS 2788-1986, a többnyelvű teauruszokra az ISO/DP 5954-1978 szabvány vonatkozik

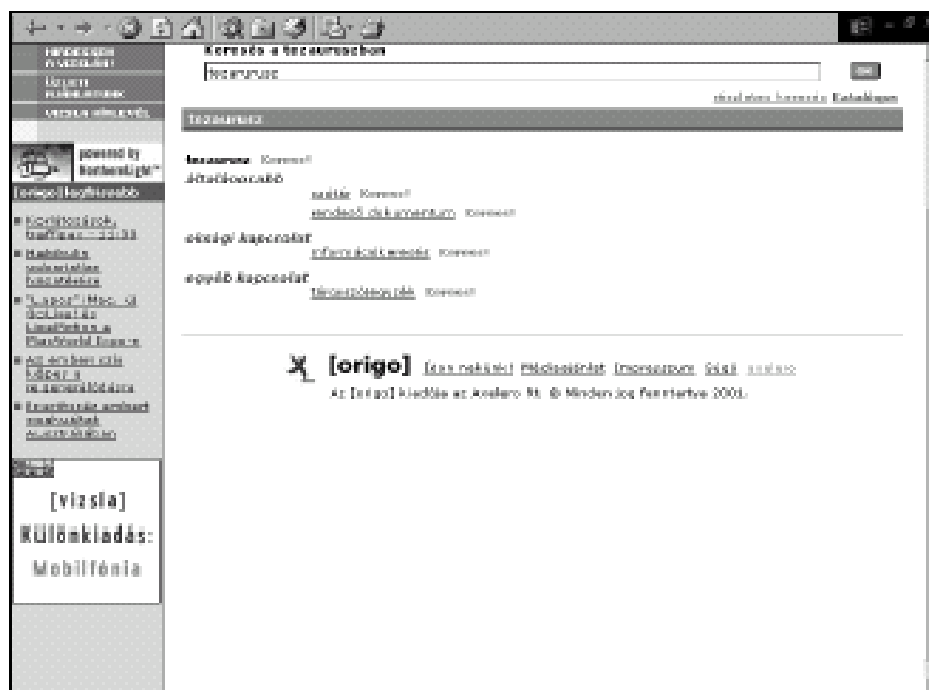
hogy nem találunk semmiféle kapcsolatot a tárgyszavak között, vagy csak nagyon kevés kapcsolat szerepel a tárgyszójegyzékekben, és ezek is csak “lásd” utalások.

A kapcsolatok – akár csak az általános magyarázatok (M) és használatra vonatkozó megjegyzések (H) – a keresőt támogatják a legalkalmasabb keresőszó kiválasztásában. Tezauruszok esetében általában részletesebb tájékoztatásra számíthatunk, mint tárgyszórendszerek esetében. **A továbbiakban a tezauruszokkal foglalkozunk, a mondottak értelemszerűen alkalmazhatók a tárgyszórendszerekre is.**

A gyakorlati alkalmazás során a tezaurusz szabványos kapcsolati jelei helyett (magyar tezauruszokban általában az F/A, R/E, X stb.) egyes esetekben a felhasználó által könnyebben értelmezhető jelöléseket szoktak alkalmazni (általánosabb/speciálisabb, oksági kapcsolat, egyéb kapcsolat stb.).

A 23. ábrán az AltaVizsla keresőszolgáltatás megjelenített tezauruszcikke látható.

*



23. ábra Tezaurusz-cikk az AltaVizsla internetes keresőszolgáltatásban. Bármelyik aláhúzott kapcsolódó kifejezésre rákattintva annak a tezauruszcikke jelenik meg (böngészés a tezauruszon belül). A “Keress!” parancsra kattintva a szóban forgó kifejezéssel kapott találatok jeleníthetők meg (keresés az adatbázisban). Webhelye: <http://vizsla.origo.hu/thesaurus?t=tezaurusz>

*

A fenti példában felülettel vezérelt tezauruszhasználatot láhattunk. Számos távoli hozzáférésű adatbázisban paranccsal vezérelt tezauruszhasználat is megvalósítható.

A DIALOG által szolgáltatott MEDLARS rendszerben például megadva az **E (arthroplasty, replacement)**

parancsot megjelenik a zárójelben szereplő kifejezés (jelentése: csontpótlás, beültetés) teljes tezauruszcikke (a vezérszó részletes adatai a 25. ábrán láthatók).

E1	arthroplasty, replacement
E2	BT arthroplasty
E3	NT arthroplasty, replacement, hip
E4	RT joint prosthesis

Az “E E3” paranccsal megkapjuk az “arthroplasty, replacement, hip” tezauruszcikkét (a félkövéren szedett E parancs a keresőszó tezauruszcikkének megjelenítésére utasít).

Egyes rendszerekben megjeleníthető a tezaurusz hierarchikus része is. A 4. ábrán az OSZK-tezaurusz/Köztaurusz taxauruszának hierarchiája látható. Az Egyesült Államok orvostudományi

szakkönyvtárának (National Library of Medicine, NLM) tezauruszböngészőjében (MeSH Browser, <<http://www.ncbi.nlm.nih.gov/entrez/meshbrowser.cgi>>) megadva a “arthroplasty, replacement” kifejezést, megjelenik a keresett szó hierarchiája

All MeSH Categories

Analytical, Diagnostic and Therapeutic Techniques and Equipment Category

Surgical Procedures, Operative

Orthopedic Procedures

Arthroplasty

Arthroplasty, Replacement

Arthroplasty, Replacement, Hip

Arthroplasty, Replacement, Knee

All MeSH Categories

Analytical, Diagnostic and Therapeutic Techniques and Equipment Category

Surgical Procedures, Operative

Prosthesis Implantation

Arthroplasty, Replacement

Arthroplasty, Replacement, Hip

Arthroplasty, Replacement, Knee

7.2.2.2 A “lásd”, a “lásd és” és a “lásd vagy”-utalás szerepe a keresésben

A legfontosabb kapcsolatok a nemdeszkriptorok és deszkriptorok közötti “lásd” utalások. A nemdeszkriptorok helyett a deszkriptorokat kell használni a keresésre (a szinonimákkal és kváziszinonimákról részletesebben a 2.2 fejezetben

foglalkoztunk). Meg kell azonban említeni, hogy egyre nagyobb azoknak a keresőrendszereknek a száma, melyekben a nemdeszkriptor is megadható keresőszóként; a rendszer automatikusan a deszkriptorral (is) osztályozott tételeket keresi meg.

Az előbb már említett NLM (National Library of Medicine) PubMed adatbázisában (<<http://www4.ncbi.nlm.nih.gov/PubMed/>>) keresve megadjuk például a “prosthesis implantation” (protézisbeültetés) keresőszót, mire megjelenik a találati oldal, melyen a “Details” (keresési kép részletezése) paranccsal megjeleníthetjük azt a keresőképet, mellyel a rendszer a megadott keresőszó alapján ténylegesen keresett.

(("prosthesis implantation"[MeSH Terms] OR "prostheses and implants"[MeSH Terms]) OR prosthesis[Text Word])

Ebből kiderül, hogy a rendszer automatikusan bevonta a nemdeszkriptorokat is a keresésbe (a keresőnek tehát nem kellett tudnia arról, hogy a nemdeszkriptornak mi a deszkriptora, és fordítva).

Vannak olyan rendszerek, melyekben a nemdeszkriptorokkal is lehet osztályozni. Az ilyen rendszerekben kereséskor megadható, hogy csak azokat a találatokat kérjük, melyeket a megadott nemdeszkriptorral osztályoztak, vagy a rendszer használja az automatikus “lásd” utalást. Ezt nevezik opcionális “lásd” kapcsolatnak.

Ennek főleg olyankor vannak előnyei, ha a nemdeszkriptor a deszkriptor valamelyik fajtája. Ha például egy rendszerben úgy döntöttek, hogy a felsőfokú igazgatással foglalkozó dokumentumok az “államigazgatás” deszkriptorral osztályozandók, akkor létrehozták az alábbi kapcsolatot (a teljes tezauruszcikk a 3. ábrán látható):

felsőfokú igazgatás

L államigazgatás

Előfordulhat, hogy a dokumentumban a felsőfokú igazgatás fogalmát használják, és ennek a kereső számára van jelentősége. Ha az adott rendszerben létezik opcionális “lásd” kapcsolat, akkor a dokumentumokat a példa esetében mindig a nemdeszkriptorral osztályozzák. Amikor olyan kereső

kérdezi le a rendszert, aki szeretné azokat a találatokat megkapni, melyek dokumentumaiban a “felsőfokú igazgatás” fogalma alapján tárgyalják a tartalmat, kérheti, hogy a rendszer csak azokat a találatokat adja ki, melyeket a “felsőfokú igazgatás” nemdeszkriptorral osztályoztak. Azok a keresők viszont, akik ezt nem kérik, bármelyik keresőszót is adják meg, megkapják mind az “államigazgatás”, mind a “felsőfokú igazgatás” ismérvekkel ellátott találatokat.

Egyes – elsősorban az interneten működő – keresőrendszerekben a tezauszban szereplő szinonimák (nemdeszkriptorok) csak külön kérésre vonhatók be a keresésbe. A “psychcrawler” keresőszolgáltatásban például a @ műveleti jelet megadva a keresőszó után a rendszer automatikusan elvégzi a keresést a szinonimák (nemdeszkriptorok) szerint is. Azért adtuk meg a zárójelben a nemdeszkriptor kifejezést, mert többnyire nemcsak a szinonimák – melyekből kevés van –, hanem az adott rendszerben annak tekintett ún. “kváziszinonimák” bevonásáról van szó; a szinonimák és kváziszinonimák egyaránt a nemdeszkriptorok szerepét játsszák. Az “államigazgatás” esetében az előbbi példában megadott “felsőfokú igazgatás” nem szinonima, hanem kváziszinonima.

A “Psychcrawlerben” a

satellite@

keresőkép(részlet) azt jelenti, hogy a keresésbe bevonják a “satellite” keresőszó tezauszban szereplő szinonimáit is.

A “lásd ÉS” (szabványos jele L&) kapcsolat esetén a nemdeszkriptort egyszerre két vagy legfeljebb három deszkriptor helyettesíti az adott rendszerben.

Például a

vadászkutya

L& kutya
vadászat

esetén a “vadászkutya” helyett a

“kutya” ÉS “vadászat”

keresőképet kell megadni.

A “lásd VAGY” (szabványos jele LV) kapcsolat esetén a nemdeszkriptor helyett a kapcsolatban megadott kifejezések közül kell kiválasztani a kereső számára relevánsat. Például az

igazgatás

LV államigazgatás
közigazgatás
vezetés

esetén az indexelőnek és a keresőnek el kell döntenie, hogy a három vagylagosan használandó kifejezés közül melyiket vagy melyeket válassza. A fenti nemdeszkriptorcikk alapján többféle keresés is elképzelhető. Például

a kereső egyet talált:

“államigazgatás”

a keresőnek mindegyik releváns:

“államigazgatás” VAGY

“közigazgatás” VAGY

“vezetés”

a kereső pontosít:

“államigazgatás” VAGY

“közigazgatás” ÉS NEM

“vezetés”

a kereső biztosra akar menni, csak azokat a dokumentumokat akarja megtalálni, melyekben a szerzők az állam- és a közigazgatás kérdését egy dokumentumon belül mindkét fogalomkör szerint tárgyalják:

“államigazgatás” ÉS

“közigazgatás”

7.2.2.3 A generikus (faj–nem) és a partitív (rész–egész) kapcsolat szerepe a keresésben

A keresésben kitüntetett szerepet játszik a generikus (faj–nem) kapcsolat, melyet a magyar teauruszokban az A, illetve az F relációjellel jelölnek (angolul NTG [narrower term generic], illetve BTG [broader term generic]).

Ha nem találunk elég információt a keresőszó alapján, célszerű átfogóbb, általánosabb jelentésű fölérendeltjével (az F relációval megadott nem-fogalmával) keresni. (Például az “agár” esetén a “kutya”, a “felületi keménység” esetében a “keménység” kifejezéssel). Fordítva: ha túl sok a találat, érdemes az alárendelt kifejezésekkel szűkíteni a keresést. Ezt “bővítő keresésnek” (broadering search) vagy “szűkítő keresésnek” (narrowing search) nevezik.

Elvileg arra is lehetőség van, hogy az alárendeltek vagy fölérendeltek automatikusan bevonják a keresésbe (upposting).

A MEDLINE azon kevés adatbázis közé tartozik, ahol ez lehetséges: a keresőképbe bevont teaurusz-kifejezések után írt felkiáltójellel nemcsak az első szintű alárendeltek lehet bevonni a keresésbe, hanem az összes alárendeltet. Például a szívbetegséget jelentő “heart diseases” deszkriptornak a MEDLINE teauruszában egész sor fajtája szerepel. Megadva az alábbi keresőképet

S heart diseases!

minden, a teauruszban szereplő szívbetegségfajtaival osztályozott dokumentumtétel egyszerre “beszippantható”. A felkiáltójelet ebben az összefüggésben “fogalmi műveleti jelnek” (concept operator) nevezik.

A generikus relációhoz hasonló, hierarchikusan szűkítő, illetve bővítő szerepet játszanak a keresésben a partitív reláció szerint végzett keresések. Itt azonban csak arra számíthatunk, hogy az egésztest jelentő keresőszóval talált dokumentumtételek halmaza valószínűsíthetően nagyobb lesz, mint a részt jelentő keresőszóval talált tételeké, de ez nem biztos. Másrészt sokkal kevésbé számíthatunk arra, hogy a két esetben lényegében hasonló tartalmú találatokat kapunk. A “tanuló” például része az “iskolának”, de ebből nem következik, hogy valamelyiknek szükségképpen nagyobb az irodalma, és a két fajta irodalom csak érintkezik egymással. A partitív reláció bevonásának a keresésbe mégis haszna lehet, mert megújíthatja, innovatívabbá teheti a kutatást. A 3. ábrán látható teauruszcikk esetén például az “államigazgatás” alapján keresve fölvetődhet, hogy a vele “egész-kapcsolatban” álló (T relációjellel jelölt) “államjog” ugyancsak használható találatokat hoz a konyhára. Ezt a relációt használva a kereséskor kellő kritikával kell kezelni a találatokat.

7.2.2.4 Az oksági, rendeltetési stb. és az egyéb rokonsági kapcsolatok szerepe a keresésben

A magyar szabványban az E/R (eszköz rendeltetés), illetve az X (rokonsági) jelekkel jelölt kapcsolatok elsősorban arra valók, hogy a kereső figyelmét fölhívják: létezhet teljesen más keresési szempont is, mint az, amelyet eredetileg kiválasztott. A “kés” rendeltetése a vágás, de ebből nem következik, hogy a késről, illetve a vágásról szóló dokumentumoknak tartalmilag köze volna egymáshoz. Mégis lehetséges, hogy a keresőnek új ötleteket adnak.

A 3. ábrán látható teauruszcikk esetében például az államigazgatás kérdéseit kutató kereső az R relációjellel jelölt (“következménye, tárgya”) “közjó” deszkriptorral is kereshet, miáltal az államigazgatás erkölcsi aspektusaira utaló találatokat kaphat.

7.2.2.5 A korábban/későbbben használt deszkriptorokra utaló kapcsolatok

Ezeket a kapcsolatokat a már régen működő és teauruszt használó információkereső rendszerekben a gyakorlat mintegy “kikövetelte” (a szabványok ugyanis még ma sem tartalmazzák). Arról van szó, hogy a tartalmi feldolgozás során egyes ismérvek helyet (melyek a teauruszban lexikai egységek) idővel mások használata vált szükségessé. Az adatbázisban sokszor nem lehet vagy nem célszerű az érintett dokumentumképekben, melyek a régi ismérvet tartalmazzák ezeket a régi ismérveket kicserélni az újakra. Vagy azért, mert a rendszer nincs ilyen cserékre fölkészítve (ez ma már ritkább

eset), vagy mert az adatbázis korábbi állományai például CD-ROM-on forgalomban vannak, a régebbi kiadású lemezek adatbázisait még használják azok, akik megvették.

Ilyenkor olyan relációjelölésekkel, mint például “Korábban” vagy “Előbb” – a MEDLARS-ban például “Previous indexing” – adják meg az előzőleg használt deskriptor(oka)t, egyes rendszerekben még a használat időszakát is megjelölve (lásd a 25. ábrát).

7.2.2.6 A megjegyzések és magyarázatok szerepe

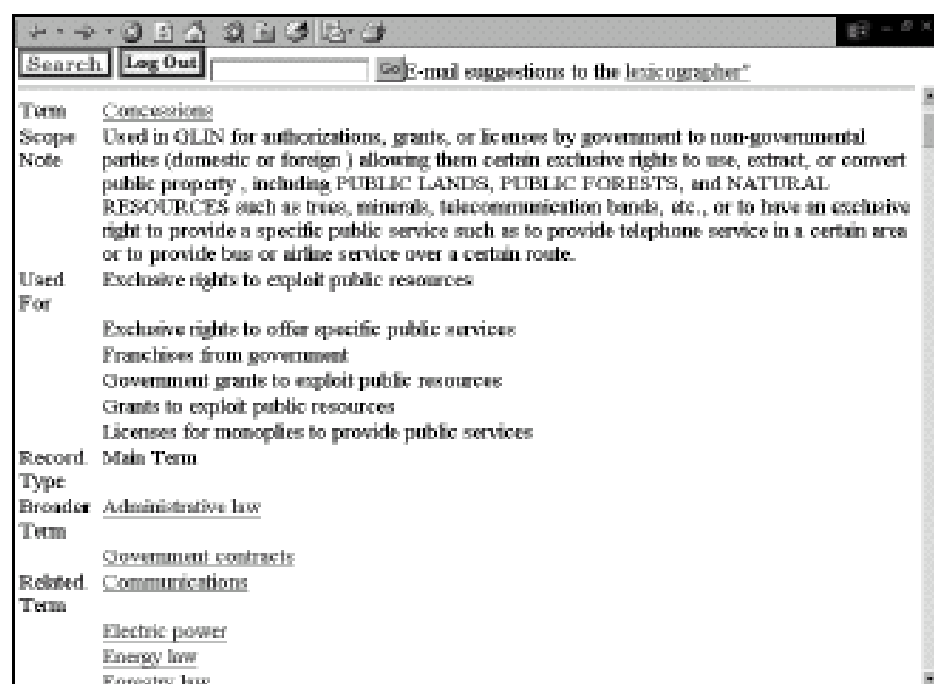
A teauruszokat használó nagyobb, jelentősebb információkereső rendszerekben a lexikai egységekhez a felhasználási tapasztalatok alapján egyre több megjegyzés és magyarázat kapcsolódik. Ezek vonatkozhatnak a lexikai egység meghatározására, értelmezésére, használatára (használatát szabályozó utasítások), használatának történetére, belső, teaurusz-szerkesztési kérdésekre, a lexikai egység forrására, kapcsolatára osztályozási rendszerekkel, a lexikai egység tárgyköri besorolására (szakcsoportra) stb.

E megjegyzések a kereső számára is gazdag tárházai az információknak, melyek alapján keresési stratégiáját tökéletesítheti, mivel felvilágosítanak arról is, hogy az adott rendszerben a választott keresőkifejezéseket hogyan használják a tartalmi feldolgozás során. A 3. ábrán az OSZK-teaurusz/Köztaurusz teauruszcikkében láthatunk példát a megjegyzések használatára. Ennél még részletesebb megjegyzésekre is bőven van példa.

Az Egyesült Államok Kongresszusi Könyvtára webes keresésekhez készült teauruszainak honlapjáról (<<http://www.pmei.com/lexico.html>>) elérhetjük például az átfogó jogi teauruszt (Global legal information network thesaurus). Rákattintva a teaurusz nevére, megjelenik a “lexicographernek” nevezett keresőprogram beviteli oldala. Megadva mondjuk a “concessions” (koncesszió) keresőszót, a 24. ábrán megjelenik a keresett deskriptor cikke.

A deskriptorcikkben a “koncesszió” (“Concessions”) vezérszó értelmezését tartalmazza a “Magyarázat” (“Scope Note”), használatának szempontjait az “Használati megjegyzés” (“Used For”) mező. A fölérrendelt (“Broader Term”) vagy rokon (“Related Term”) kapcsolatokra kattintva a kapcsolódó deskriptorok teauruszcikke jeleníthető meg.

*



24. ábra. Teauruszcikk megjegyzéseinek és magyarázatainak példája. Webhelye (ahonnan több lépésben elérhető): <http://www.pmei.com/lexico.html>

*

A National Library of Medicine honlapján (<www.nlm.nih.gov/>) megadva a könyvtári szolgáltatásokat (“Library services”), majd az orvostudományi teauruszt (Medical Subject Heading, MeSH), azon belül pedig az online keresést, “online searching”) megjelenik a MEDLINE

tezaurusz böngészőjének (MeSH Browser) felülete. Itt megadva a keresett kifejezést (példánkban “arthroplasty, replacement”), megjelennek a lexikai egység összes adatai (25. ábra).

Az “arthroplasty, replacement” (csontpótlás, beültetés) vezérszóra vonatkozó részletes használati megjegyzéseket az “annotáció” (“Annotation”) mező tartalmazza. A megjegyzés (“Scope Note”) mezőben látható a meghatározás.

A különféle alakváltozatokat képviselő belépőszavakat az “Entry Term” mezők tartalmazzák. Az “Allowable Qualifiers” mezőben láthatók azoknak az altárgyszavaknak a jelzetei, melyekkel a deszkriptor finomítható. Ha például összekapcsoljuk a deszkriptort a CT (“contraindications”, ellenjavaslatok) altárgyszóval, akkor nem általában a csontpótlásról van szó, hanem annak ellenjavallatairól.

*

MeSH Heading	Arthroplast, Replacement
Tree Number	E04.555.110.110
Tree Number	E04.650.110
Annotation	ARTHROPLASTY, REPLACEMENT, HIP & ARTHROPLASTY, REPLACEMENT, KNEE are also available; do not use [any] except by MeSH definition; DF: ARTHROPLASTY REPLACE
Scope Note	Partial or total replacement of a joint.
Entry Term	Joint Prosthesis Implantation
Entry Term	Arthroplasties, Replacement
Entry Term	Replacement Arthroplasties
Entry Term	Replacement Arthroplasty
See Also	Joint Prosthesis
Allowable Qualifiers	AE CL CT EC HI IS MO MP NU PGRH SN STTD UT VE
Entry Version	ARTHROPLASTY REPLACE
Previous Indexing	Joint Prosthesis (1974-1997)
Previous Indexing	Prosthesis and Implants (1966-1970)
Previous Indexing	specific joint (1966-1970)
History Note	MS, use JOINT PROSTHESES 1979-97
Unique ID	D019643

25. ábra. Deszkriptor adatlapja az Egyesült Államok orvostudományi szakkönyvtárának (National Library of Medicine) tezauruszából. Webhelye (ahonnan több lépésben elérhető): www.nlm.nih.gov/

*

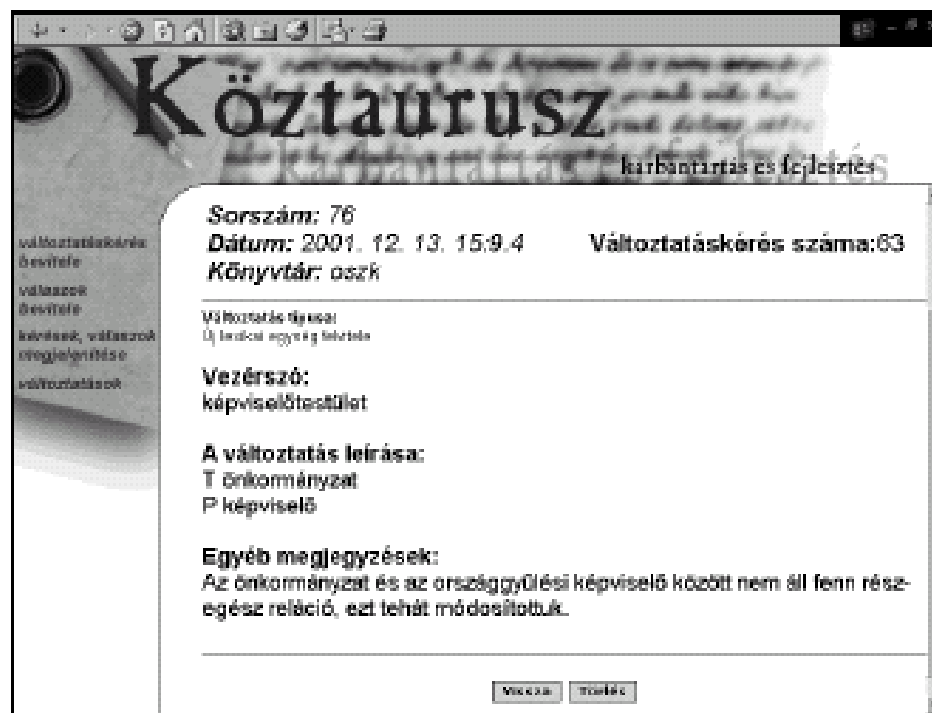
7.2.2.7 A tezauruszok karbantartása

Akárcsak az egyetemes osztályozási rendszerek (TO, ETO), a nagy tezauruszok is bonyolult, a használat során módosításokra, bővítésekre szoruló, bonyolult rendszerek. A változásokról jó tudnia a keresőnek annak érdekében, hogy a kiválasztott rendszert a legjobban használhassa fel. A karbantartást az információkeresés világában talpon maradt tezauruszok szerkesztősége végzi – valójában éppen azért váltak be a nemzetközileg is ismert tezauruszok, mert önálló szerkesztőségek gondjaira bízták őket.

Általában nemcsak rendszeresen frissítik az online tezauruszt, hanem gyakran nyomtatott változatot is évente közreadnak. A MEDLINE tezauruszát például minden évben kiadják. Az Országos Széchényi Könyvtár és a közművelődési könyvtárak közös tezauruszának is van nyomtatott változata. A 3. ábra bal oldalán látható két tezauruszcikk például ennek a tezaurusznak az internetről is letölthető, nyomtatható változatából származik³⁰. Az interneten elérhető közös karbantartó felületén a “változtatások” paranccsal megjeleníthető az utolsó változat (és kiadás) óta bekövetkezett módosítások kumulált jegyzéke, a “kérdések, válaszok megjelenítése” paranccsal pedig megállapíthatók a változtatások. A 26. ábrán az új lexikai egység fölvételének értesítőlapja látható.

³⁰ OSZK-tezaurusz/Köztaurusz. Az Országos Széchényi Könyvtár és a közművelődési könyvtárak átfogó tezaurusza: 2.0 változat / Főszerkesztő Ungváry Rudolf; [közr. a] Magyar Könyvtárosok Egyesülete és az Országos Széchényi Könyvtár. – Budapest: MKE, OSZK, 2001. < <http://www.oszk.hu/ujdonsag/tezauruj.html> >

*



26. ábra Az OSZK-tezaurusz/Köztaurusz új lexikai egységének fölvételéről tájékoztató lap.
Webhelye: <http://mvkkvar.hu/kozt/>

7.2.3 Keresés osztályozási rendszer (ETO és mások) alapján

7.2.3.1 A hierarchikus rendszerek (és az ETO) szükségessége

Az osztályozási rendszerek a keresés klasszikus eszközei. Az osztályokat a 2.6 fejezetben ismertetett hierarchikus szerkezet jellemzi, a kereső e hierarchiában böngészve mintegy “becserkészi” a számára releváns információkat.

Nagy előnye ennek a keresési lehetőségnek, hogy az egyre szűkebb tartalmú, pontosabban meghatározott tárgykörök felé lehet haladni, miközben megőrizhetjük a rálátásunkat a rendszer egészére. A rendszer ismerője mindig tudhatja, éppen “hol van” a hierarchikus “világban”, és könnyen dönthet arról, mennyire akar speciálisabb információkhoz jutni.

A tárgyszavakkal és a deszkriptorokkal/nemdeszkriptorokkal végzett keresésre az jellemző, hogy a kiválasztott keresőszavak révén nagyon pontosan irányított a keresés. A kereső azt, és csakis azt az információt láthatja a keresés eredményeként, amelyre vonatkozó kérdését megfogalmazta. Ezekkel az eszközökkel keresve általánosságban azt mondhatjuk, hogy viszonylag nagy a keresés pontossága, de ezzel arányosan nagy a veszteség. Ezek az eszközök a nyelvtől függenek, melyet nagyon különböző módon lehet használni, amin az információkereső nyelvek szabványosítása sem képes lényegében változtatni.

A hierarchikus rendszerekben végzett böngészéskor a kereső mindig láthatja az általa megcélzott tárgykör teljes környezetét. Ezáltal nemcsak egyszerű és gyors helyesbítéseket hajthat végre a keresésben, hanem minimalizálhatja a veszteséget. Egész általánosan azt mondhatjuk, hogy a hierarchikus eszközökkel végzett kereséskor kisebb a pontosság, de lényegesen nagyobb a teljesség.

Valójában mindig két keresőrendszerrel célszerű keresni ugyanazt a témát, mert amit az egyik eszközzel nem lehet megtalálni, többnyire megtalálható a másikkal és fordítva. E tapasztalatnak tulajdonítható, hogy a világ könyvtáraiban a tezauruszok megjelenésével párhuzamosan egyáltalán nem szorult vissza a két legjelentősebb hierarchikus osztályozási

rendszer, az angol nyelvterületen elterjedt Dewey-féle³¹ Tizedes Osztályozás és a még nála is elterjedtebb (és Magyarországon nagy hagyományokkal rendelkező) Egyetemes Tizedes Osztályozás (ETO, angol rövidítéssel UDC). Ellenkezőleg: megjelentek már ezeknek a rendszereknek a webes változatai, például az amerikai könyvtárakat bibliográfiai rekordokkal ellátó On-line Computer Library Center (OCLC, a világ legnagyobb közös katalógizálási rendszere) által forgalmazott WebDewey.

Tévhit, hogy a számítógéppel támogatott információkeresés világában ezek az osztályozási rendszerek kevésbé hasznosíthatók. Ellenkezőleg: éppen a gépesített keresés teszi lehetővé, hogy a hierarchikus rendszerek összes tulajdonságai kihasználhatók legyenek. **Éppen ezért a gépesítés fejlődésével szembehelyezkedő megoldás, ha háttérbe szorítanak vagy éppen elsorvasztanak például az ETO-val végzett osztályozást a tárgyszavas és a teauruszokkal végzett tartalmi feldolgozás javára. Mindkettőre szükség van.**

Az más lapra tartozik, hogy korábban az információkereső rendszerek jelentős részét még nem készítették föl igazán a hierarchikus osztályozási rendszerek lehetőségeit kihasználó keresésekre – amiben éppen az említett tévhit képviselői voltak a ludasak. A helyzet mára már részben megváltozott. Egyre több, az interneten is hozzáférhető katalógusban lehet már a TO, illetve az ETO szerint keresni.

7.2.3.2 A hierarchikus rendszerek előnyei a számítógépes keresésben Átfogó és egyetemes jelleg

Az említett két egyetemes osztályozási rendszer az ismeretek teljességét fogja át, a tárgykörök szinte teljes választékát nyújtják. Kétségtelenül bonyolult szerkezetű rendszerek, emiatt hosszú jelzetek reprezentálják a specifikus fogalmakat. A hierarchikus jelzetek csonkolásával éppen a gépi keresés biztosíthatja a legáttekinthetőbb hozzáférést ezekhez az általánosabb fogalmakhoz.

Ha például valaki nem talál elég információt a varsákkal és csapdákkal végezhető halfogásról, melynek jelzete a 639.2.081.16, akkor a 639.2.081 jelzet alapján a halászati eljárásokról, a 639.2 rövidítés alapján a halászatról általában, és a 639 rövidítés alapján a haltenyésztésről, vadászatról általában tájékozódhat a jelzet jobb oldalán végzett csonkolásával.

A számítógépes rendszer a hierarchikus osztályozási rendszerek alapján végezhető keresést más vonatkozásban is sokkal rugalmasabbá teheti. Megoldható – ami cédulakatalógusok esetében még kilátástalannak tűnt –, hogy keresni lehessen például az összetett ETO-jelzetek egyes elemeivel, az általánosan közös alosztások túlnyomó és a korlátozottan közös alosztások számottevő részével. Bár kétségtelen, hogy erre a Magyarországon ma működő integrált rendszerek túlnyomó többsége (még) nem nyújt lehetőséget.

Az önállóan kereshető – például földrajzi és történelmi – jelzetelemeket az arra felkészített rendszerekben szabadon lehet kombinálni mind a főtablázati számokkal, mind egymással. Nagyobb online katalógusokban e jelzetrészek alapján az információk páratlan tematikai gazdagsága bontakozik ki a kereső előtt. Az alábbiakban két részletet láthatunk az OSZK online ETO-mutatójából. Az első esetben a mutatót az idő szerinti közös alosztás jelzetei szerint permutálva rendezték (dőlt szöveggel a jelzetekkel jelölt osztályok tartalmát adjuk meg).

A felső idézőjelek az idő szerinti alosztást, felosztási szempontot jelenti, a (439) Magyarország jelzete (földrajzi alosztása), az emelt idézőjelek között az idő szerinti alosztások szerepelnek.

930.85 (439)	".../08"	Magyarország művelődéstörténete a 9. századig
930.85 (439)	".../08" (02.053.2)	Magyarország művelődéstörténete a 9. századig, ifjúsági művek

³¹ Az 1851-ben született Melville Louis Kossuth Dewey második és harmadik keresztnévét a születése idején egyesült államokbeli előadókörútját tartó Kossuth Lajos tiszteletére kapta.

091.14 (439)	".../1526": 016 1526	előtti Magyarországi kódexek bibliográfiája
27(439)	".../16"	magyarországi keresztény egyháztörténet a 17. századig
355.48 (439)	".../17" (075.3)	Magyarország hadtörténelme a 18. századig, középiskolai tankönyv
338 (091) (439)	"09"	Történelmi források Magyarország gazdasági helyzetéről a 10. században
903/904 (439)	"14"	Régészeti leletek, régiségek a 15. századi Magyarország idejéből
903/904 (439)	"14" (075.8)	A 15. századi Magyarország régészete, egyetemi tankönyv
34 (439)	"15" (093.2)	A 16. századi magyarországi jog történelmi forrásai oklevelei
348.1/.7 (439)	"15" (093.2)	A 16. századi Magyarország katolikus egyházjogának forrásai, oklevelei

A második mutatót a (4–62)-es földrajzi alosztás szerint permutálva rendezték (a kezdő és a befejező kerek zárójel azt jelzi, hogy földrajzi alosztásról, szempontról van szó, a 4 Európát, a –62 az államok szövetkezését jelenti, az ETO-ban ezzel osztályozzák az Európai Uniót).

339.923 (4–62)		Nemzetközi gazdasági együttműködés az EGK-ban
(4–62)		európai szövetséges államok, Európai Unió
327.39 (4–62)		Egyesülési mozgalmak, Európai Unió
341.17 (4–62)		a nemzetközi jog és az Európai Unió
338 (4–62)	"199"	az Európai Unió gazdasága az 1990-es években
338 (4–62)	"199" (083.41)	az Európai Unió gazdasága az 1990-es években, statisztikai táblázatok
339.5 (4–62)	"199"	az Európai Unió külkereskedelme az 1990-es években
339.727.22 (4–62)	"199"	tőkebehozatal, külföldi tőke az Európai Unióban az 1990-es években
35 (4–62)	"199"	igazgatási jog, közigazgatás az Európai Unióban az 1990-es években
338.27 (4–62)	"20"	az Európai Unió gazdaságára vonatkozó előrejelzések a 20. században

Látható, hogy az egyes jelzetrészekre való keresés olyan válogatási lehetőségek számára nyitja meg az utat, amely csak a TO és az ETO típusú osztályozási rendszerek segítségével valósítható meg – ha az információkereső rendszereket fölkészítették rá.

A fenti példánál maradva kiválogatható az Európai Unióra és elődeire, társintézményeire, az egész egységesülő Európára vonatkozó irodalom, mely együtt található meg a (4–62)-es földrajzi alosztásnál.

További példa: a “rendszerátalakítás az európai volt szocialista országokban” téma irodalma a Kelet–Európát képviselő (4–11) földrajzi alosztás és az "1989/199" idő szerinti alosztás segítségével gyűjthető össze:

(4–11) ÉS "1989/199"

Kiválogathatók adott tárgykörhöz tartozó felsőoktatási tankönyvek a (075.8) jelzetrész és a tárgykör összekapcsolásával. Például a 14. századi Magyarország hivatalos forrásait, okleveleit tartalmazó dokumentumok a bibliográfiák kivételével a következő keresőképpel:

"14" ÉS (439) ÉS (093.2) ÉS NEM 016

Kiválogathatók például a romákkal foglalkozó irodalom a (=914.99)-es, a zsidósággal kapcsolatos irodalom a (=924)-es, a határon túli magyarokra vonatkozó irodalom a (=945.11)-es etnikai alosztásnál, a rendszerváltás és a rendszerváltást követő évek irodalma a "1989/199"-es idő szerinti alosztásnál, a reneszánsz művészetet tárgyaló művek a .034-es ("pont nulla harmincnégyes) korlátozottan közös alosztásnál, a gyermekeknek szóló irodalom a (02.053.2)-nél, a gyermekekről szóló könyvek a -053.2-es (olvasása: "mínusz nullaötvenhárom pont kettes") személyek szerinti alosztásnál, a magyarra fordított szépirodalmi művek az =945.11-es nyelvi alosztásnál, az angolra fordítottak az =20-nál, a németre fordítottak az =30-nál stb.

A korlátozottan közös alosztásokkal önállóan végzett keresések több mint szépséghibája, hogy ugyanazok a korlátozottan közös alosztások az egyes főtblázati számokhoz kapcsolva mást-mást jelentenek. Ezeket az alosztásokat a kereséskor önállóan használva adott esetben zaj keletkezhet.

A mondottak nemcsak az ETO-val osztályozott nagy katalógusok páratlan forrásgazdagságát, hanem az ETO rugalmas, többoldalú alkalmazhatóságát is igazolják a számítógépes információ keretei között. Ebben ugyanis megjelenik az ETO kézi rendszerekben rejtetten maradó ama tulajdonsága, hogy a jelzetrészek valójában a természetes nyelvekre, tárgyszavas és deskriptoros rendszerekre jellemző mellérendelő (posztkoordinált) használatot is lehetővé teszik.

Noha – mint mondtuk – a könyvtári rendszerek többsége az ETO-jelzeteivel való illetően rugalmas keresésre még nincs fölkészítve, a lehetőségeket azért ismertettük, hogy ezek tudatában, értő szakmai közvélemény igényeire és nem utolsósorban nyomására válaszul a könyvtári rendszerek fejlesztése megfelelő irányban mozdulhasson tovább.

Az ETO-t a fenti rugalmassággal kezelni képes gépi információkereső rendszerben ugyanakkor radikálisan csökken a kettősponttal jelölt, előre megadott mellérendelés (a prekoordináció, az ETO "nyelvén" viszonyítás) jelentősége. Miután a számítógépes rendszerekben szabadon kombinálhatók a különféle jelzetek, nincs szükség rá, hogy előre összekössék őket a kettőspontos viszonyítási jelzettel. A kézi szakkatalógusokban például az alábbi témát csak viszonyítással lehetett megoldani:

621.643.24 : 669.37 : 696.121: 726.5

Rézből készült csőelvezető csatornák
építészetileg helyes kialakítása templomokban

A kettőspontos osztályozást ma már nagyrészt nem is használják.

Online katalógusokban szükségtelen kettősponttal összekapcsolni a jelzeteiket; elég külön-külön megadni őket a dokumentumképben, mert az információkereső rendszer logikai műveleteivel a jelzetek a keresőképben összerendezhetők.

Az ilyen rendszerekben már csak akkor használják a viszonyítást, ha a fogalmat csak két szám kombinációjával lehet az ETO-ban kifejezni. Például a "kulturpolitika" esetében: 323 : 008.

Böngészhetőség

Nagy segítség, hogy ezekben a rendszerekben egyrészt a hierarchikus szerkezet (lásd az 1. és a 2. ábrán látható hierarchiákat), másrészt a fentiekben ismertetett permutált mutatók jóvoltából szinte kézenfogva vezetnek el elsősorban a járatlanabb keresőt a kívánt információkhoz.

Az alábbi ábrán egy elektronikus könyvtár Tizedes Osztályozási rendszeren alapuló webkatalógusának böngészhető részlete látható.



27. *ábra.* A “The Internet Public Library” hierarchikus webkatalógusának részlete a Tizedes Osztályozás jelzeteivel Webhelye: <<http://www.iporg/reading/books/index.html>>

Ismertség és nemzetközi jelleg

A TO-t vagy az ETO-t szinte minden könyvtárban ismerik, többnyire helyben is használják, és ezért a könyvtárakban föl vannak készítve a keresés támogatására. Mivel a számjelzetek következtében nem függenek adott nyelvtől, a világon mindenütt használhatók; aki az egyik országban már eligazodott bennük, másik országban, más nyelvterületen ugyanúgy használni tudja őket.

Értelmezhetőség

A hierarchikus osztályozási rendszerekhez már kezdettől fogva készültek természetes nyelvű mutatók. Például az alábbi osztályra

612.014.46 Kémiai anyagok és mérgek hatásai. Fizikai–kémiai hatások, például ozmózis, kolloidok, ásványi sók stb. hatása

az ETO középkiadásának mutatójában a következő mutatószavak utalnak (a releváns mutatószavakat félkövéren jelöltük):

Ásványi olaj 621.892.2

Ásványi olaj növényi vagy állati olajjal való keveréke 621.892.4

Ásványi só /állati termékek feldolgozása 637.041.046

Ásványi só hatása /fiziológia 612.014.46

...

Kémiai állandó 54.04

Kémiai analízis /diagnosztika 616–074

Kémiai anyag hatása /fiziológia 612.014.46

...

Fizikai–kémiai energia hasznosítása erőgépekben 621.499

Fizikai–kémiai finomítás /kőolajfeldolgozás 665.663

Fizikai–kémiai hatás /fiziológia 612.014.46

...

Méregbehatás /fiziológia 612.464

Méreg hatása /fiziológia 612.014.46

Méreg hatása a bélben /fiziológia 612.354

Méreg hatása a májra /fiziológia 612.354

...

Az ETO-ban tehát nem csak böngészve lehet keresni, hanem a mutatóban talált természetes nyelvű keresőszavak alapján közvetlenül is kiválaszthatjuk az adott jelzetet és rajta keresztül a találatokat. Ugyanakkor a jelzet egyértelmű értelmezéséhez és alkalmazásához követelmény, hogy szemügyre vegyük helyét a hierarchiában. Természetesen ehhez is arra van szükség, hogy az információkereső rendszer kezelni tudja az összekapcsolt ETO-jelzeteket – és erre ma még ugyancsak kevés rendszer képes

A mesterséges nyelven megfogalmazott jelzetek értelmezhetőségét néhány korszerűbb gépi információkereső rendszer azzal is segíti, hogy az adott jelzethez bármikor megjeleníthető a jelzet magyarázata és egyéb adatai, azaz megállapítható az osztály tartalma. Az előző ábrán már a jelzetek szerinti mutatóban megadták a rövid tartalmat.

8. Az információkeresés értékelése

Már a múlt század második felében foglalkozni kezdtek az információkeresés eredményeinek értékelésével.³² Az első értékelő vizsgálatra 1953-ban az Egyesült Államokban (az egyik első teaurusz használatával kapcsolatban) került sor, egy évvel később pedig Angliában Cyrill W. Cleverdon ismertette eredményeit. A matematikai logika két művelője, M. E. Maron, és J. L. Kuhn feltehetően először 1960-ban használták a relevancia fogalmát. Tefko Saracevic 1975-ben írt monográfiájában pedig megfogalmazta a “relevancia filozófiáját”. Hamarosan konferenciákra is sor került. Az első ilyen tárgyú konferenciát 1966-ban rendezték, azóta kisebb-nagyobb rendszerességgel olykor évente több információkereséssel foglalkozó nemzetközi konferencia szervezésére kerül sor.³³

Az információ relevanciája egész általánosan annak ismérve, hogy a keresés adott kérdésre milyen mértékben a kérdésnek megfelelő információkat válogatta ki, függetlenül attól, hogy a felhasználó is relevánsnak ítélte-e meg őket, tud-e valamit kezdeni vele vagy sem. Mennyiségileg a relevancia adott keresőkérdés vonatkozásában a visszakeresett tételek azon hányada, amely objektív értelemben megfelel a feltett kérdésnek.

A keresés szempontjából az adatbázisban egyrészt releváns és nem releváns, másrészt megtalált és nem megtalált információ létezik. Táblázatban:

*

	releváns	nem releváns
megtalált	A	B
nem megtalált	C	D

6. táblázat Információk a relevancia szempontjából.

*

A (megtalált releváns információk)	= találatok (hits)
B (megtalált nem releváns információk)	= zaj (noise, flas drop)
C (releváns nem talált információk)	= veszteség (misses, losses)
D (nem releváns, nem talált információk)	= érdektelen információk

A keresés tökéletes, ha a B és a C halmaz üres. Nem ennyire eszményi, de még mindig optimális, ha minden releváns információt kiadott a rendszer, de kiadott irreleváns információkat is, azaz csak a

³² Szöveggyűjtemény, 2. köt. p. 195.

³³ Az egyik legismertebb az 1963-ban indult, az illinois-i egyetemen az adatfeldolgozás könyvtári alkalmazásának intézete által évenként rendezett konferencia (Annual Meeting of the Clinic in Library Applications of Data Processing), és az 1976-ban indult, minden év végén Londonban megrendezett "Online information Meeting", melyet az Online Review, a The Electronic Library és még több más szakfolyóirat kiadója, a Learned Information (Europe) Ltd. szervez.

C halmaz üres. A tökéletesség mértékét a teljességgel (recall, retio) és a pontossággal (precision) jellemzik.

A két fogalmat először William Perry használta a szemantikai kódok használhatóságának elemzésével kapcsolatban (ő még pertinencia-tényezőknek, Cleverdon 1963-ban relevancia-tényezőknek nevezte a kettőt; mai nevüket Gerard Salton adta 1965-ben).

A keresés teljessége (recall ratio) a megtalált releváns dokumentumok aránya az összes (akár talált, akár nem) releváns dokumentumhoz viszonyítva. Azaz

$$T = \frac{\text{talált releváns dokumentumok száma}}{\text{releváns dokumentumok száma az adatbázisban}} \quad 0 \leq T \leq 1$$

1. képlet

A keresés pontossága a talált dokumentumok összességén belül a releváns dokumentumok aránya. Azaz

$$P = \frac{\text{talált releváns dokumentumok száma}}{\text{talált összes dokumentum száma}} \quad 0 \leq P \leq 1$$

2. képlet

A teljesség növelésével a pontosság elkerülhetetlenül csökken, és ez fordítva is igaz. A két tényező viszonya a keresési stratégia segítségével szabályozható (lásd a “3.6 A keresési stratégia” című fejezetet).

Az intellektuális keresések értékelésében még ma is nagy a bizonytalanság. Egy 1991-ben a DIALOG-rendszerben végzett online keresésekre vonatkozó vizsgálat a következő összegzéssel zárul:

“Az eredmények igen kevés törvényszerűséget mutattak. A szakma valószínűleg nincs is tudatában, hogy a keresések eredményeiben mennyire kevés a megegyezés... Az online keresés messze van attól, hogy tudománynak lehessen nevezni... Nincs semmiféle elfogadható módszer, átfogó és következetes irányelv, mely az online információkeresés útmutatójaként szolgálhatna. A keresés még mindig művészet, mégpedig meglehetősen kevésbé megfogható és meghatározható művészet.” Ha jobban meggondoljuk, ez nem is olyan nagy baj: a művészet segítségével a legnehezebb emberi problémák oldhatók meg. (A keresésben alkalmazható intuitív, “művészi” – pontosabban heurisztikus – módszerekkel könyvünk “3.7.2 Kereső taktikák” című fejezetében foglalkozunk.)

Az egzakt szemlélet tükrében – talán éppen a keresés intuitív jellege miatt – a keresőrendszerek értékelése különösen nehéznek bizonyult, és az értékelés általános elmélete még várat magára.

Azzal a kérdéssel, hogy mit értékeljünk, mi az, amit mérni lehet, már 1966-ban Cleverdon foglalkozott.³⁴ Munkájában hat fő mérhető mennyiséget sorolt fel:

- (1) a gyűjtemény vélhető érdeklődést lefedő volta, vagyis hogy milyen mértékben tartalmaz a rendszer releváns dokumentumokat;
- (2) az időtényező, vagyis hogy mekkora az átlagos időeltérés a keresési kérdés feltétele és a válasz megadása között;
- (3) az erőfeszítés a használó részéről annak érdekében, hogy keresőkérdésére a választ megkapja;

³⁴ A ma már klasszikusnak számító cranfieldi és aberyswythi vizsgálatokra vonatkozóan lásd az Osztályozás és információkeresés c. szöveggyűjtemény 2. kötetében az “Információkeresés értékelése” című részt (p. 195-202).

- (4) a keresés teljessége, vagyis a megtalált releváns dokumentumok aránya az összes (akár talált, akár nem) releváns dokumentumhoz viszonyítva;
- (5) a keresés pontossága, vagyis a talált dokumentumok összességén belül a releváns dokumentumok aránya.

Az első négy könnyen becsülhető. Mélni igazán a teljességet és a pontosságot kellene, e kettő fejezi ki a leginkább az információkereső rendszer hatékonyságát. Másszóval feltételezik, hogy ez a mértéke a rendszer azon képességének, hogy megtalálja a releváns dokumentumokat, s ugyanakkor visszatartja az irrelevánsakat. Cleverdon óta se szeri, se száma a legkülönbözőbb ilyen jellegű vizsgálatoknak

A baj csak az, hogy e két paraméter alapja nem kezelhető ugyanolyan egzakt módon, mint maga a két paraméter. A problémát Van Rijsbergen 1987-ben magyarul megjelent könyvében így fogalmazta meg.

“[A teljesség és a pontosság] valamilyen módon még mindig igényli a relevancia meghatározását... A relevancia [azonban] szubjektív fogalom. Különböző használók különböző álláspontot foglalhatnak el adott dokumentumnak egy kérdésre vonatkozó relevanciáját vagy irrelevanciáját illetően... a kísérleti eredményeket... rendszerint jóhiszemű használókból csalták ki, olyanoktól, akik valamilyen szakterületen dolgoznak és információk igényeik vannak... Olyan helyzetben vagyunk..., hogy nagyszámú kérdéshez ismerjük a 'helyes' válaszokat. Az információkeresés területén általános feltételezés, hogy ha egy információkereső stratégia elegendően jól működik nagyszámú kísérleti feltétel mellett, akkor vélhetően jól fog működni gyakorlati szituációban is, amikor a relevanciát nem ismerjük előre.”

Másszóval mindannak, ami objektív – a pontosság és a teljesség –, olyasmi az alapja, ami szubjektív. Ez minden értékelés alapvető ellentmondása az információkeresés világában. Egy jólnevelt természettudós vagy matematikus ezért kicsit olyan viszonyban a relevanciával, akár az ördög a szenteltvízzel.

Még súlyosabb a helyzet a pertinenciával. A relevancia azt fejezi ki, hogy mekkora a közelség a felhasználói kérdés és a talált dokumentumok tartalma között (azaz azok a dokumentumok, melyek a kérdésnek megfelelnek, relevánsak); a vele szorosan összefüggő pertinencia nem más, mint a felhasználói információszükséglet és a talált dokumentumok tartalma közötti megfelelés (tehát nem pusztán a kérdést, hanem azt, ami a “tudaton belül van”, kellene összevetni az eredménnyel). Ezért aztán a természettudományos alapon közelítő szerzőknél a pertinencia még mutatóban sem fordul elő.

Figyelemre méltó, hogy Mortimer Taube 1965-ben lándzsát tört a nyilvánvalóan szubjektív relevanciafogalom használata mellett, és élénken tiltakozott mindenfajta “matematizált” relevancia bevezetése ellen. Az utóbbin alapuló képleteket, melyekkel az információkereső rendszerek hatékonyságát számszerűen értékelhették, pszeudo-matematikai konstrukciónak tekintette.

Van Rijsbergen beszámol arról, hogy elvileg ugyan van olyan relevanciafogalom, amelyet objektívnek tekinthetünk, s amelyet “logikai relevanciának” nevezhetünk.

“...ez a relevanciafogalom jelenleg nagyon korlátozottan használható csupán. Ennek fő oka, hogy olyanfajta rendszert, amely ahhoz kellene, hogy olyan információkereső stratégiát alkalmazzunk, amely csak a logikailag releváns dokumentumokat keresi vissza, még nem hoztak létre... Lehetséges, hogy egy ilyen típusú rendszer mérete túlzott a mai számítógépek számára; de a végső szót majd a jövő mondja ki.”

Azóta, hogy a fenti a sorokat leírták, több mint tíz év telt el, feltehetően túl rövid idő ama bizonyos jövőnk bekövetkeztéig.

Ma az elmúlt évtizedek vizsgálatai alapján nagyjából a következő általános tapasztalatok szűrődtek le az információkeresés értékelésével kapcsolatban:

- a legjobb eredmények a teljesség vonatkozásában az olyan dokumentációs nyelvekkel érhetők el, melyekben a nem összetett szavak az uralkodók;
- a keresési szabályok befolyásolják a teljesség és pontosság alakulását;

- a szabadon választott szavakkal végzett keresés soha sem volt olyan rossz, mint a legrosszabb szabványosított nyelven (a legrosszabb tárgyszójegyzékkal, a tezaurusszal, a legrosszabbul használt ETO-val) végzett keresés, és soha sem volt olyan jó, mint a legjobb szabványosított nyelven végzett keresés (másszóval ha a szabványosított nyelv jó, illetve az ETO-t jól használják, akkor jobb eredmény érhető vele el, mintha szabadon választott szavakkal végezzük a keresést, ha viszont a szabványosított nyelv rossz, illetve az ETO-t rosszul használják, akkor a szabadon választott szavakkal végzett keresés a jobb);
- a specifikus keresés nagyobb teljességet eredményez, mint a kevésbé specifikus keresés.

9. Az információkereső gondolkodás története

9.1 Az információkeresés fogalmának megszületése

“Az emberi agy nem így működik. Asszociációkat követ.

Megragad valamit és már kapcsol is tovább...”

Vannevar Bush: As we may think [Ahogy gondolkodhatnánk]

In: Atlantic Monthly, 1945 július.³⁵

Az információkeresés szakterületének keresztapja az eredetileg bölcész végzettségű Mortimer Taube (1910–1965), aki egyik 1950-ben írott tanulmányában használta először az angol vadászati nyelvben honos “retrieval” kifejezést: “...searching and retrieval of information from storage encoding to specification by subject.” [...a tárgy pontos meghatározása érdekében kódolt információ keresése, becserkészése és visszanyerése a tárolóból]. Magyarul a leginkább még a “becserkészés és visszanyerés” összetétel áll a legközelebb a szó eredeti angol jelentéséhez. Információkeresésen (retrieval, information retrieval) a kereséssel összefüggő teljes folyamatot értjük, kezdve a keresőkérdés elemzésével, a keresőprofil és a keresési stratégia és taktika kialakításán és a tárolóban végzett keresőműveleteken át a találatok képzéséig és kiadásáig.

Az “információkeresés” kifejezést ma két értelemben használják a szakirodalomban.

Jelenti egyrészt az információk intellektuális, programrendszerek (főleg adatbázis-kezelő rendszerek) segítségével végzett keresését. Könyvünkben ezzel foglalkozunk könyvtári nézőpontból.

Jelenti másrészt az automatikus indexelési és osztályozási eljárásokat, melyekben természetes nyelvű szövegek gépi feldolgozásával tárják föl a dokumentum tartalmát.³⁶

Mind az első, mind a második értelemben vett eljárás esetében felvetődik, hogy milyen információkról van szó. A problémára rávilágít Cornelis van Rijsbergen kissé egyoldalú értelmezése:

“Ami azt illeti, sok esetben az információkeresés megfelelő módon leírható úgy is, hogy az ‘információ’ szót egyszerűen a ‘dokumentum’ szóval helyettesítjük. ...a legtökéletesebben szókimondó definíciót Lancaster adta meg: ‘Az információkeresés olyan általánosan elfogadott, de némileg pontatlanul használt kifejezés, amelyet a könyvemben tárgyalt tevékenység megjelölésére alkalmaztak. Egy információkereső rendszer nem ad információt a használónak kérdése témájáról, azaz nem változtatja meg ismereteit. Csupán arról ad tájékoztatást, hogy van-e, vagy nincs, s ha van, hol van olyan dokumentum, amely választ ad kérdésére.’ Ez a meghatározás kizárja... a kérdés-

³⁵ <<http://www.theAtlantic.com/atlantic/atlweb/flashbks/computer/bushf.htm>> Magyarul: Út az új gondolkodás felé.

In: Hypertext+Multimédia. Oktatási segédanyag. A szöveget vál. Sugár János; szerk. Klaniczay Júlia. – Budapest:

Artpool, 1996. p. 3–14. Továbbá: Ahogy gondolkodhatnánk. In: Komenczi Bertalan; J. C. R. Licklider, a

katedrálisépítő. In: Neumann Jánostól az Internetig. Akik nyomot hagytak a 20. századon, 4. Napvilág Kiadó, Bp. 1999.

³⁶ Az automatikus eljárások meghatározására a “Keresési stratégia” című fejezet elején röviden már kitértünk. A kérdést részletesebben tárgyalja: Van Rijsbergen: Információ visszakeresés. [közr. az] Országos Széchényi Könyvtár Könyvtártudományi és Módszertani Központ. – Budapest: Múzsák Közművelődési Kiadó, 1987. Az alábbi idézet helye: p. 7. Angol webhelye: <<http://www.dcs.glasgow.ac.uk/Keith/Preface.html>>

felelet rendszereket. Ugyancsak kizárja azokat az adatkereső rendszereket, amelyek például a tőzsdei árfolyamok online szolgáltatására valók.”

Van Rijsbergen az elsődleges tényadatok keresésére nem az információ-, hanem az adatkeresés kifejezést javasolja. Valójában nem igaz, hogy a hivatkozási (másodlagos) adatokat/információkat szolgáltató információkereső rendszer nem változtatja meg használójának ismereteit; az is megváltoztatja, csak éppen nem az elsődleges, hanem a másodlagos információk (tehát a hivatkozások) szintjén. Hiszen a használat előtt nem tudta (vagy rosszul tudta), milyen dokumentumot és hol keressen, utána – szerencsés esetben – viszont tudni fogja.

A információkeresés kontra adatkeresés terminológiával pedig az a baj, hogy a tényadatok/faktografikus adatok keresői és az ilyen rendszerek kezelői nem feltétlenül tudnak erről a finom distinkcióról, és a maguk rendszereit nem adatkereső, hanem – joggal – információkereső rendszereknek nevezik. Ezért helyesebb, ha az információkereső rendszert ebből a szempontból mind a másodlagos (a dokumentumtétel, reprezentációk, hivatkozások), mind az elsődleges információk keresőrendszerének generikus fölérendelt fogalmának tekintjük, és ha szükséges pontosítani, faktografikus, illetve bibliográfiai vagy hivatkozási információkeresésről beszélünk. Kötetünkben e “hivatkozási” információk keresésével foglalkoztunk.

9.2 Eseménytörténet

A szakirodalom mennyiségének növekedésével párhuzamosan a 20. század első évtizedeiben egyre nyilvánvalóbbá vált, hogy mind a Dewey-féle Tizedes Osztályozás, mind utóbb az Egyetemes Tizedes Osztályozás, mind pedig a hagyományos – Cutter-féle – tárgyszavas osztályozási eljárások túlságosan nehézkesen használhatók, ha már nemcsak könyvekről, hanem a rendkívül speciális tartalmi feldolgozást igénylő folyóiratcikkekről van szó. A tárgyszavas rendszerekről – mivel természetes nyelven alapultak – a dokumentátorok jelentős részének volt olyan az érzése, hogy elvileg alkalmasabb lehetnének a finomabb és gyorsabb keresés céljaira, mint az ETO és a hozzá hasonló mesterséges nyelven alapuló rendszerek. Ösztönös várakozásukat idővel kiábrándulás követte, mivel a tárgyszavas rendszereket a könyvtárosok továbbra is azoknak a hierarchikus szerkezeti elveknek megfelelően fejlesztették tovább, ahogy azt még a 19. század második felében Cutter elkezdte. Ezek a könyvek keresésére készült fő-, al- és melléktárgyszavas tárgyi (tárgyszavas) katalógusok³⁷ a részletes és gyors információk szolgáltatására törekvő szakemberek szemében a harmincas–negyvenes évekre az információkeresésre való alkalmatlanság jelképeivé váltak.

A második világháború után felgyorsult a könyvtári és információs szolgáltatások fejlődése, az információkeresés iránt is növekedni kezdtek a követelmények. Az e szakterületen “dokumentációnak” nevezett tevékenységi és ismeretkör fokozatosan átalakult, kibővült információtudománnyá (tájékoztatótudománnyá). Ebben az időszakban alkotta meg Claude Shannon és Warren Weaver a matematikai információelméletet, Norbert Wiener a kibernetikát, kezdték építeni a kereskedelembe forgalmazott nagyszámítógépeket.

Fontos szerepet játszott a fejlődés élénkítésében, hogy 1950-ben hozzáfogtak az Egyesült Államokban a tudományszervező Vannevar Bush kezdeményezésére a National Science Foundation égisze alatt a nem-konvencionális információs rendszerek fejlesztéséhez.

Bush a háború alatt körülbelül hatezer vezető amerikai szakember munkáját koordinálta annak érdekében, hogy a tudományos kutatásokat a hadviselés szolgálatában felhasználhassák; ő irányította az atombomba előállításának tudományos programját. Még a háború befejezése előtt, 1945 áprilisában tanulmányt írt “As we may think”³⁸ címmel, melyben az emberi észjárásról

³⁷ Részletesebben lásd az “Osztályozás és információkeresés” c. szöveggyűjtemény 1. kötetében, p. 46.

³⁸ As we may think. In: Atlantic Monthly, 1945, July, No. 176, p. 101–108). Magyarul: Út az új gondolkodás felé. In: Hypertext + multimédia. Oktatási segédanyag. A szöveget vállal. ...Sugár János; szerk. Klaniczay Júlia. Budapest, Artpool, 1996. p. 3–14.

közelebb álló, a mai hipertext-kapcsolódásokra emlékeztető világméretű információkereső rendszer vízióját fogalmazta meg, összekapcsolt számítógépek hálózatával, melyekben az analógia és az asszociációk útján működő rendszerekkel végzik a keresést. Ezzel mintegy előrevetítette egyrészt a szemantikai kapcsolatokon (s végső soron a gondolati asszociációkon)s, velük a strukturált információkereső nyelvi szótárakon (a teauruszokon), másrészt az internet hipertext-csatolói alapuló keresést.

Az információkereső rendszerek fejlődését meghatározóan támogatták az ebben az időszakban alakult védelmi szervezetek és más hadiiparilag fontos intézmények is. Mindezzel szoros szellemi és anyagi kölcsönhatásban született meg a 20. század 50-es éveiben az információkeresés (information retrieval) szakterülete is.

A praktikus keresőszolgáltatások kialakítása terén úttörő szerepet játszott az információkeresés fogalmának megalkotója, Mortimer Taube. Az általa útnak indított koordinált indexelés³⁹ kiszabadította az indexelést (a “mutatókészítést”) a hagyományos, hierarchikusan strukturált tárgyszavas eljárás szemléletéből és ezzel az információkereső gondolkodás szellemét is kiszabadította a palackból. Ugyanakkor felhívta a figyelmet az indexelés/osztályozás nyelvi aspektusaira, a szintaktikai és szemantikai problémákra.

A felhasználás tapasztalatai alapján hamar világossá vált: a kereséskor figyelembe kell venni elvont szemantikai (analitikus) összefüggéseket (mi minek a fajtája, része, oka, tulajdonsága stb.), hogy az eljárás hatékony legyen. Fred Jonker fogalmazta meg 1959-ben, hogy az egyes szavak specifikus, konkrét kiválasztása valójában csak az egyik valószínű választás bekövetkeztét jelenti a lehetséges szóválasztások spektrumából. Ennek alapján alkotta meg Jonker az információkeresés egyik klasszikus fogalmát, a deskriptív kontinuumot, amely szerint az információkereső nyelv kifejezései, melyekkel a dokumentumok tartalma leírható, végtelenül finomítható, megszakítatlan, folyamatos értékvilágot alkotnak, akár csak a fizikában a kontinuumként (nem részecskeszerkezetűként) vizsgált anyag.⁴⁰

Mindennek nyomán kezdtek szerkeszteni az 50-es évek végétől a gépi információkeresés céljaira strukturált információkereső–nyelvi szótárakat, a teauruszokat.

A koordinált indexelés egy másik, az információk kereshetőségét javító felismerést is inspirált, mely első lépésben a különféle számítógépekkel előállítható mutatókban (indexekben) öltött testet.⁴¹ Az eredetileg textilmérnök Hans Peter Luhn 1956–58-ban vezette be a folyóiratcikkek címszavait eredeti szövegkörnyezetükön belül kereshetővé tevő permutált indexet⁴², s ennek nyomán vált általánosan használttá a kulcsszó fogalma a dokumentumok szövegében a keresés alapjául szolgáló szavakra. (A “kulcs” itt azonosítót jelent, melyen keresztül a dokumentum elérhető.)

A gépi indexekkel párhuzamosan kezdték használni az ötvenes évek második felétől az információkereső rendszer fogalmát, és terjedt el – igazából még a számítógépek számottevő megjelenése előtt, az ötvenes évek végétől, hatvanas évek elejétől, még a manuális rendszerekben – a rekord, a fájl, az elérés, az invertálás és az adathordozó, az indexelő/információkereső nyelv, az indexelés és a dokumentumkép/keresőkép fogalma.

A 60-as évek elején jelentek meg a második generációs számítógépek és a nagy tárolóképességű mágnesszalagos információhordozók. Az adatokat a kulcsszavakat tartalmazó invertált gépi indexek segítségével már közvetlenül, s ezáltal technikailag gyorsan el lehetett érni. Ez azt jelentette, hogy az állományokat nem kellett minden rendszerben a tárolóban folyamatosan végigolvasni ahhoz, hogy adott kulcsszóval keresett adatot elérjenek, hanem elég volt a gépi indexben megkeresni a kulcsszót, és onnan közvetlenül “ugrani” lehetett a keresett adatra (például bibliográfiai rekordra).

³⁹ Szöveggyűjtemény, 2. köt. p. 31.

⁴⁰ Szöveggyűjtemény, 2. köt., p. 91., 138.

⁴¹ Szöveggyűjtemény, 1. köt. p. 351., 2. köt. 151.

⁴² Szöveggyűjtemény, 2. köt. p. 69., 135. Először 1953-ban a CIA (az amerikai központi hírszerző ügynökség) dokumentátorai kezdték gépi segítséggel permutálni a címszavakat.

Ezáltal az adatbázis bármely invertált adattétele a mutatószaván (címezésén) keresztül közvetlenül elérhetővé vált.

Mai szemmel nézve rendkívül nehézkes volt azonban a keresés. A felhasználó nem vehetett részt a feldolgozásban. Az általa meghatározott keresőkérdést (pontosabban a keresőprofil) a számítógép üzemeltetője vette át, futtatta a programot, amely elvégezte a keresést, és az eredményeket visszaadta. Ezt nevezték a számítástechnikában általában kötegelt (batch) feldolgozásnak. Elsősorban az ún. szelektív információterjesztésre (SDI, selective information dissemination) volt alkalmas a felhasználók érdeklődési körét kifejező keresőkérdéseket kötegelt feldolgozásban rendszeresen feldolgozták és az eredményeket a felhasználóknak eljuttatták. Az adatbázisokban a felhasználó nem kereshetett közvetlenül.

A 20. század második felében egyre nagyobb szükség lett az információkeresés kérdéseinek összegezésére, a keresőrendszerek átfogó, általános tárgyalására. Az 1960–90 közötti időszakban sorra jelentek meg a mára klasszikus kézikönyvek és monográfiák. Az elért eredmények összefoglalásában az egyik legtermékenyebb szerző az angol Brian C. Vickery (1918). Harold Borkoval (1922), Douglas John Foskettel (1918–1999) és Jesse Hauk Sherával (1903–1982) együtt az ún. “második angol tanárnemzedék” nemzetközi hírű képviselői, Samuel C. Bradforddal (1878–1948) és Mortimer Taubeval (1910–1965) együtt a könyvtári információkereső gondolkodás szakmai úttörői.

A Kongresszusi Könyvtárban helyezték üzembe 1960-ban az első könyvtári célú számítógépet. Az évtized közepére működni kezdett számos nagy gépi információkereső rendszer (például a MEDLARS, az ERIC).

A 70-es évek elején jelentek meg a harmadik generációs számítógépek, a közvetlen elérésű mágneslemezes háttértárolók, melyekben lehetővé vált, hogy a lemez bármely fizikai részét mindig ugyanannyi idő alatt érjék el. Ez rendkívül lerövidítette a keresés idejét. Ettől kezdve a feldolgozásban a végfelhasználó is közvetlenül részt vehetett. Megszületett az interaktív (párbeszédés) munkakapcsolat, melyben a gép kezelője által bevitt parancsok és kérdések, meg a gépi válaszok váltakoznak, s ezáltal afféle “társalgás” játszódik le a gép és a felhasználó között. Erre alapozva alakult ki a helyi online információkeresés

Helyi online információkereséskor ugyanabban a szervezeti egységben, ahol az adatbázis van, egyetlen rendszer információkereső eszközeivel, helyileg végzik a keresést. Ilyen, amikor valaki a könyvtáron belül használja a könyvtár nyilvános online katalógusát. Az optikai lemezek megjelenésével lehetővé vált teljes adatbázis tartalmának rögzítése egyetlen, szállítható hordozón, a helyi online információkeresés fogalmába az ebben való keresés is bekerült. Mivel a CD-ROM adatbázisok száma rendkívül nagy, és szervezettségük a lehető legváltozatosabb, speciális hozzáértés szükséges ahhoz, hogy valaki a különféle CD-ROM adatbázisok telepítéséhez és használatához professzionálisan értsen. Erre a feladatra, valamint a helyi és távoli adatbázisokban való online keresésre született meg idővel az elektronikus könyvtárosok szakterülete.

1966-ban a Kongresszusi Könyvtárban elkészült az első mágnesszalagos adatsere-formátumot (MARC). Ennek nem csak abból a szempontból volt korszakos jelentősége, hogy a bibliográfiai állományokat a könyvtárak egymástól át tudták venni, hanem azt is jelentette, hogy a keresés mind inkább, és egységesen eme adatsere-formátum meghatározott mezőjéhez kapcsolódott.

Az adatbázisokban az adatszerkezetet kezdetben hierarchikusan, később hálósan, mára pedig már relációsan szervezték. A hierarchikus szervezés nyomán alakult ki a mező–almező szerkezet. Mivel a könyvtári adatsere–formátumokat a hierarchikus adatbáziskezelés idején kezdték kialakítani, szerkezetüket a mai napig ez a szervezési mód határozza meg, holott azóta a relációs adatszerkezetek megjelenésével az adatbáziskezelés maga teljesen megváltozott. Az információkeresés ezért ma is meghatározott mezőkön belüli almezők tartalmaira irányul.

A 60-as évek végén már folytak az első információs hálózatok tervezési munkái, 1970-ben működni kezdett az első információs célok megoldására hivatott távközlési hálózat, a TYMNET, 1972-ben a DIALOG Information Services, később egész sor további kereskedelmi adatbázis-szolgáltató révén már távolból is online elérhető lett számtalan nagyobb és kisebb adatbázis (például az OCLC, a MEDLARS (MEDLINE), az ERIC adatbázisa). Ezt nevezték távoli hozzáférésű keresésnek.

Elkészültek az első online teauruszok (mind az adatbázisok, mind a teauruszok száma mára több százra növekedett)..

A távoli hozzáférésű kereséskor során egy vagy több, nem a közvetlen szomszédságban lévő adatbázishoz férhetünk hozzá a távközlési hálózat és megfelelő adatátviteli rendszer segítségével. A keresőnek ilyenkor nagyon eltérő információkereső-nyelvi eszközöket kell ismernie ahhoz, hogy a különböző szervezettségű adatbázisokban kiismerje magát. Erre a keresési feladatra idővel információközvetítők szakosodtak; mára kialakult az információs brókerek szakterülete.

Az online rendszerekben a felhasználónak nem kell ismernie magát a keresőrendszert, ahogy a személyautó vezetőjének sem kell értenie a jármű szerkezetéhez. A felhasználó egyre inkább olyan felületen keresztül végezhet az interaktív keresést, melynek megjelenési formája egyre közelített a megszokott, mindennapi gondolkodáshoz és jelhasználathoz. A felhasználóbarát végfelület fejlődéséhez döntő lökést adott a hipertext megjelenése. A nevet először Theodor Holm Nelson használta 1965-ben. 1967–68 között Andries van Dam dolgozta ki az első hipertext szerkesztő rendszert, de még közel két évtizednek kell eltelnie, hogy 1987-ben a piacon megjelenjen az első kereskedelmi forgalomban kapható hipertext szoftver.

Ezzel párhuzamosan született meg a távolsági adatátvitel céljaira az első TCP/IP szabvány, s ennek nyomán az Internet. 1989-ben pedig Tim Berners-Lee a genfi Európai Részecskefizikai Laboratóriumban (CERN) Nelson inspirációjára előterjesztette a World Wide Web tervét. A 90-es évektől viharos gyorsasággal terjedt el a globális hipertext-alapú internet navigációs rendszer, a web. A 90-es évek közepétől működni kezdtek az első webes indexelő és kereső szolgáltatások (mint az AltaVista, a Yahoo!, a magyar Heuréka, Hudir és AltaVizsla).

Ugyancsak a 90-es évek elejére az addigra hagyományossá vált hierarchikus adatbázis-kezelő rendszerekkel kezelt könyvtári rendszereket (például az ISIS-t) fokozatosan felváltották azok a könyvtári rendszerek, melyek már relációs adatbáziskezelőket használnak (például az ORACLE), és amelyek webes végfelhasználói keresőfelülettel is rendelkeznek.

A 21. század fordulójára teljesen átalakult az információkeresés technikai környezete. Ebből azonban hiba volna azt következtetni, hogy a lényege is megváltozott volna. A lényege ugyanis, hogy az információkeresést végső soron mindig az ember végzi, akit még a legtökéletesebb gépesítéssel és keresőeszközökkel sem lehet kiiktatni anélkül, hogy az egész folyamat értelmét ne vesztené.

A könyvtári információkeresés “könyvtári” jelzője ezt a lényeget jelzi: a keresés elvégzését és a keresők támogatását hozzáértő szakemberekkel.

Irodalom

Az alábbiakban csak a Magyarországon elérhető néhány fontos könyvre utalunk. Az információkeresés tengernyi nemzetközi dokumentuma közül mutatóban csak egyetlen emelünk ki, hogy az olvasó annak bibliográfiája alapján tovább léphessen. A magyar dokumentumok közül az első tételként megadott szöveggyűjteményen kívül kettőt vettünk fel a rövid irodalomjegyzékbe. Egyebekben a szöveggyűjtemény tartalmára, valamint a szöveg lábjegyzeteiben adott hivatkozásokra utalunk. A forrásokat fordított időrendi sorrendben, rövid annotációval ellátva közöljük.

Ungváry Rudolf, Orbán Éva: Osztályozás és információkeresés. Kommentált szöveggyűjtemény. [közread az] Országos Széchényi Könyvtár. – Budapest: OSZK, 2001.

1. köt., Az osztályozás és elmélete. – 543 p.

2. köt., Az információkeresés és elmélete. – 535 p.

Az információkereső nyelvekről, az információkeresésről és az internet használatáról neves klasszikus és mai külföldi szerzők műveiből hosszabb szemelvények találhatóak, melyekre könyvünkben a releváns helyeken utalunk is.

Pálvölgyi Mihály: Információfeldolgozás, információkereső nyelvek. – Szombathely, Berzsenyi Dániel Főiskola, 1998. – 327 p.

Az információkereső nyelvek és használatuk részletes ismertetése mellett a könyv számos tezaurusz példáját is közli.

Hartley, R. J., Large, Andrew, Tedd, Lucy A.: Information seeking in the online age. Principles and practice. – London. Brower–Saur, 1998. – 308 p.

Az online információkeresés három egyetemi előadója által írt, a gyakorlat által áthatott szemléletű könyvben az elméleti részek mellett részletes parancsnyelvi keresési példák, számtalan stratégiai és taktikai fogások találhatóak.

Az online szakirodalmi információkeresés kézikönyve. Szerk. Roboz Péter. [közread. az] Országos Műszaki Központ és Könyvtár, Építéstudományi Tájékoztató Intézet. – Budapest, OMIKK, 1985.

1. rész., – 226 p.

2. rész., – 550 p.

Rövid, ma is érvényes gondolatokat tartalmazó bevezető után az egyes távoli hozzáférésű adatbázisok tulajdonságainak, parancsnyelvének stb. ismertetését tartalmazza, helyenként ma már nem érvényes adatokkal, példái azonban a keresési technikákat jól szemléltetik.

Ungváry Rudolf: Az osztályozás alapjai: Bevezetés az információkereső nyelvek elméletébe. – Budapest: OSZK KMK, 1982. – 116 p.

A tankönyvben megtalálható az információkeresés és az információkereső nyelve való fordítás részletes leírása, ismerteti információkereső nyelvek alakjánát, mondattanát és jelentésstanát, különös tekintettel az automatizált adatbázisokban való alkalmazásukra.