

Ungváry Rudolf – Vajda Erik

Országos Széchényi Könyvtár Országgyűlési Könyvtár

Az információkeresés szavai

Az információkereséshez használhatók osztályozási jelzetek, tárgyszavak, szabadon választott szavak, tezaurusz deskriptorai, egységesített besorolási adatok, de akár ISBN vagy ISSN, országnevek kódszavai, nem könyvtári körülmények között pedig telefonszám, személyi szám, tartalomjegyzék szövege, könyv mutatószava stb. Mindezek információkereső nyelvi szavak, melyek egyben a dokumentumleíró nyelv szavai (a dokumentum ismérvei) is. A szabadon választott szavakat hibásan szabad tárgyszavaknak szokták nevezni, holott ezek – ellentétben a tárgyszavakkal – nincsenek szabványosítva, ezért nem lehetnek tárgyszavak. A velük végzett keresés a szabad szavas keresés, szemben például a kötött szavas kereséssel, mely tárgyszavakkal, deskriptorokkal, egységesített besorolási adatokkal, jelzetekkel stb. végezhető. A dokumentum szempontjából megkülönböztetjük a csak ismerv céljára meghatározott mezőkben (pl. jelzet-, tárgyszó-, deskriptor-, ISBN-mezőkben) végzett ismervmezős keresést a dokumentumot szövegesen leíró mezőkben (pl. a címben vagy a tartalmi kivonatban, referátumban) vagy a teljes dokumentumszövegben végzett szövegmezős kereséstől.

Bevezető

Az információkeresés¹ egyik lényeges, de nem kizárólagos feltétele, hogy (a) szavakat használunk hozzá. A keresés további feltétele, hogy (b) a dokumentumokat megfelelő szavak jellemezzék (lévén, hogy csak olyan szavakkal kereshetünk értelmesen, mely szavak a dokumentumokat is jellemzik), és (c) annak ismerete, hogy a dokumentum leírásán (a továbbiakban általában: dokumentumtételen) belül hol helyezkednek el ezek a szavak (a címben, a tartalmi kivonatban vagy külön erre használt mezőkben stb., vagy éppenséggel nem a dokumentum leírásában, hanem a dokumentum szövegében, illetve annak egy részében). Mindezek felől nézve a kereséseknek több változata lehet, s ezeket a változatokat meg is kell tudni nevezni.

A következőkben a fenti nézőpontból kiindulva rendszerezve tárgyaljuk a különféle keresési változatokat, nem utolsósorban azért, hogy megnevezésük egységes használatához hozzájáruljunk. Tesszük ezt különösen azért, mert az ETO (és egyéb könyvtári osztályozási rendszerek) jelzetei („szavai”) mellett (sőt igen gyakran helyett) mára általánossá vált a természetes nyelvű² kifejezések használata a tárgyköri feldolgozáskor (leíráskor) és a kereséskor, és ezzel összefüggésben gyakran és hibásan használják a „tárgyszó” kifejezést.

A keresőszó

A keresés során összehasonlítás játszódik le a kereséshez használt és a dokumentumot jellemző szavak között. A találatok annak eredményében keletkeznek, hogy van-e (és adott esetben milyen mérvű) megfelelés e – kétfajta – szó között.³

Ha például a keresést a dokumentumok címe alapján (a dokumentumtétel cím mezőjében) végezzük, és a „Magyarország” meg a csonkolt „ebtenyészt” szavakat használjuk fel a kereséshez (azaz a keresőrendszerben megfelelő műveletekkel relevánssá tehetők a szavak részei, illetve ragozott alakjai, vagy az azokból képzett szavak is, nem csak a teljes, „eredeti” szó, példánkban az „ebtenyészt”), akkor találat az a dokumentum, amelynek „A magyarországi ebtenyésztés története”, vagy „A Magyar Ebtenyésztők Egyesületének története” a címe.⁴

A felhasznált szavak státusa alapján mindig keresőszavakról van szó. Innen nézve a kereséshez használt tezaurusz lexikai egysége és a tárgyszó-jegyzékek tárgyszava, – *horribile dictu* – az ETO keresésre használt jelzete, vagy akár a szabályozott információkereső nyelvi „szótár” használata nélkül „kitalált” (helyesebben: megtalált) természetes nyelvi szó egyaránt a keresőszó szerepét játszhatja.

Azt a szót tehát, amelyet a kereséshez fölhasználunk, **keresőszónak** nevezzük.

A szavak fajtái a nyelvek szempontjából

Az információkereső nyelvek szavai

A keresőszavak elválaszthatatlanok az információkereső nyelvtől. Minden keresőszó valamilyen információkereső nyelv szava. Ezeket a nyelveket szabványosíthatják, de a keresés céljára használnak nem szabványosított nyelveket is. A szabványosított nyelvek jelentős részét szótárba foglalják (kötött szótáras nyelvek), de léteznek nem kötött szótárba foglalt szabványosított információkereső nyelvek, sőt olyan információkereső nyelvek is, melyeket nem szabványosítottak. A fontosabb eseteket az 1. táblázat tartalmazza. Ebben a nyelveket *szótári jellemvonásaik (szerkezetük) alapján* csoportosítottuk.

A nem szabványosított információkereső nyelvek csak természetes (beszélt, írott) nyelven alapulhatnak. A szabványosított nyelveken belül a strukturált nyelvek közül a szemantikai információkereső nyelvek általában természetes nyelven alapulnak (elvileg elképzelhető, hogy meghatározott fogalmak szimbólumai között fejeznek ki szemantikai relációkat). A hierarchikus nyelvek, valamint az enumeratív nyelvek a gyakorlatban alapulhatnak természetes vagy mesterséges nyelven egyaránt. Az ETO például mesterséges nyelven alapuló hierarchikus nyelv, egy dokumentum tartalomjegyzékének kifejezése pedig természetes nyelven alapszik; az országnevek kódszáva, a cégek adószáma vagy a telefonszám mesterséges nyelven, az egyszerű szójegyzék szava pedig természetes nyelven alapszik.

Az előbbieken szótári jellemvonásaik alapján tárgyaltuk ezeket a nyelveket. Egy másik rendezési szempont a *felhasználásuk* jellege. Ennek alapján a fenti információkereső nyelvek lehetnek olyanok, melyeknek szavai önmagukban egy-egy besorolási „területet”, halmazt, osztályt képviselnek. Ezek az osztályozó „nyelvek”; mivel a használatuk elvileg nem „nyelvszerű”, nem leíró-megfogalmazó, hanem besoroló, ezért *osztályozási rendszereknek* nevezik őket. Ide tartozik a legtöbb hierarchikus nyelv (például az ETO, az SZJ-számok rendszere, a szabadalmi osztályozási rendszer). Az osztályozási rendszer lehet enumeratív is (li-

neáris osztályozási rendszereknek is nevezik őket), ha nem rendelik szavait (az osztályokat) egymás fölé/alá. Például ilyen volt kezdetben a Kongresszusi Könyvtár osztályozási rendszere (amely mára már jelentős mértékben bővült hierarchikus összefüggésekkel).

1. táblázat

Az információkereső nyelvek fajtái a nyelvek szótári jellemvonásai (szerkezetük) alapján

1. szabványosított információkereső nyelv szava
 - 1.1 kötött szótárba foglalt információkereső nyelv szava
 - 1.1.1 strukturált (valamilyen rendszerbe szervezett) információkereső nyelv szava
 - 1.1.1.1 szemantikailag (különbféle értelmi és morfológiai [együttesen paradigmatis] összefüggések szerint explicite kifejezett) strukturált információkereső nyelv szava (pl. tezausz deszkriptora⁵, a „lásd” vagy a „lásd még” vagy mindkét kapcsolatot („utalást”⁶) tartalmazó tárgyszó-rendszer tárgyszava);
 - 1.1.1.2 hierarchikusan (pontosabban: többnyire nem szorosan vett, explicite kifejezett szemantikai, hanem praktikus alá-fölé rendelési viszonyok szerint) strukturált információkereső nyelv szava (pl. ETO-jelzet, NSZO [Nemzetközi Szabadalmi Osztályozás] jelzete, SZJ [szolgáltatások jegyzéke] szám, FEOR [Foglalkozások Egységes Osztályozási Rendszere]⁷, könyv „több-szintes” tartalomjegyzékének kifejezése);
 - 1.1.1.3 enumeratív (felsoroló, csak mechanikusan rendezett) információkereső nyelv szava (pl. betűrendezett, de kapcsolatokat nem tartalmazó tárgyszójegyzék szava, az országok és a nyelvek nevének szabványosított két-, illetve három karakterből álló kódszavai, könyv tárgymutatójának szava, adónem kódszáva)
 - 1.2 kötött szótárba nem foglalt szabványosított információkereső nyelv szava (pl. egységesített személy- és testületi név, egységesített cím, ISBN és ISSN, személyi szám, URL [Uniform Resource Locator, internetes dokumentumok azonosítója], cégek adószáma, magánszemélyek adóazonosító jele)
 2. nem szabványosított információkereső nyelv szava: kötetlenül, szabadon választott keresőszó, melyre legfeljebb az a szabály (de az automatizált eljárások terjedésével előbb-utóbb talán még ez se) vonatkozik, hogy ne ragozott forma legyen

Az „osztályozási rendszer” és a „hierarchikus információkereső nyelv” kifejezések tehát nem szinonimák. Nem minden osztályozási rendszer hierarchikus, és nem minden hierarchikus szerkezetű

nyelv osztályozási rendszer. Elképzelhető például olyan teaurusz, melyben a relációk közül csak a generikus (nem faj típusú) relációkat (kapcsolatokat, összefüggéseket) tüntetik föl, de ettől ez a teaurusz még nem a fenti értelemben vett osztályozási rendszer, noha hierarchikus. A következő bekezdésben ugyancsak láthatunk példákat hierarchikus rendszerekre, amelyek nem osztályozási rendszerek.

Az enumeratív nyelvek mesterséges nyelven alapuló fajtáinak kötött szótáras változatai általában osztályozási rendszerek (pl. az országok neveinek vagy az adónemeknek a kódszavai), a kötött szótár nélküli változatai (pl. cégek adószámai, a magánszemélyek adóazonosító jelei, a személyi számok) pedig soha nem osztályozási rendszerek. Az utóbbiak nem osztályt, hanem egyetlen előfordulást, dolgot jelölnek, azaz individuumok (egyedi előfordulások) kódszavai. Nem osztályt, hanem egyedi neveket jelenthetnek hierarchikus felépítésű kifejezések is. A telefonszám például hierarchikus szerkezetű. A 36 1 2243738 tartalmazza a két-karakteres országhívó számot, a körzetszámot (Budapest körzetszáma 1, az ország többi részén a körzetszámok kétjegyűek), és utánuk következik az előfizető kapcsolási száma, de mind az egyes összetevői, mind a teljes szám mindig csak egy adott dolgot azonosítanak (a példában a 36 Magyarországot, az 1 Budapestet, a 36 1 2243738 pedig a budapesti előfizetőt), szemben például az ETO-jelzettel, amely egynél több dolgot azonosíthat. Ilyen hierarchikus szerkezetű egyedi azonosítók még az ISBN, az URL, a postai irányítószámok (pl. H 1026), a cégek adószámai stb.

Ugyancsak nem osztályokat, hanem egyedeket azonosítanak az egységesített nevek és címek, valamint a nem általános földrajzi nevek. Még a csillagködök és galaxisok nevei is, hiszen a csillagköd meg a galaxis neve egy meghatározott csillagködöt vagy galaxist azonosít, és nem többet. Ennek semmi köze ahhoz, hogy egy galaxis rendkívül sok elemből – bolygórendszerekből, egyedi csillagokból, üstökösökből, meteoritokból stb. – áll, mert a galaxis neve nem az egyes összetevőket azonosítja (ezeknek adott esetben önálló azonosítója van), hanem a galaxist magát, mint egyedet. Hasonló a helyzet a(z egységesített) családnévvel. A Gundel önmagában csak a család egészét, tehát családot azonosít (voltaképpen akkor is, ha az éttermet azonosítja). Ha a család tagjaira alkalmazzuk, akkor persze már osztálynévként használjuk, de akkor nem is lehet vele az egyes családtagokat egyedileg azonosítani, mivel a Gundel név

eredeti rendeltetése is csak az, hogy magát a családot azonosítsa egy az egyben (és egységesített névként csak ebben a szerepében használható). Az itt felsorolt információkereső nyelvi szavakat éppen ezért (elsődleges) azonosítóknak nevezik (velük rendeltetésszerűen nem lehet egynél több egyedet jellemezni), az általuk alkotott nyelv egyedi vagy (elsődleges) *azonosító nyelv (rendszer)*. Az elsődleges azonosítók különleges esetei a könyvek tartalomjegyzékének egységei. Ezek ugyanis szigorúan (formálisan) véve meghatározott könyvrészeket azonosítanak mint egységeket. Értelemszerűen (lazán véve) persze e részek különféle tartalmait.

Az előbbiekkal szemben az osztályozási rendszerek jelzeteit vagy a tárgyszavakként, deskriptorokként használt ún. általános neveket (pl. kutya, galaxis) megkülönböztetésül az előbbihez *másodlagos azonosítóknak* nevezik. Ide tartoznak a könyvek mutatói is. Amikor elsődleges azonosítókkal osztályoznak dokumentumokat, akkor ugyan egy azonosító terjedelmébe több dokumentum is eshet, de ezek mind ugyanarról az egyedről szólnak, szemben a másodlagos azonosítókkal, melyek terjedelmébe különböző egyedekről szóló dokumentumok tartozhatnak. Például a „Budapest” földrajzi névvel osztályozott dokumentumok halmaza csupa azonos egyedről (kizárólag Budapestről vagy többek között Budapestről is) szóló dokumentumot tartalmaz, ezzel szemben a „kutya” tárgyszóval osztályozott dokumentumok halmaza a legritkább esetben tartalmaz csak egyetlen meghatározott kutyáról szóló dokumentumokat.

Az információkereső nyelvek elvileg lehetnek olyanok, melyek – önmagukban is egy-egy osztályt képviselő (leíró) – szavaiból össze kell (lehet) állítani azt az osztályt, amelybe a dokumentumot besorolják. A szakirodalomban elsősorban ezeket nevezik *információkereső nyelveknek*, velük mintegy leírjuk, megfogalmazzuk a dokumentum tartalmát, és ez a leírás lesz végül maga az osztály, melybe a dokumentum „beletartozik” (szemben az egyes információkereső nyelvi szavakkal, amelyek – bár önmagukban maguk is egy-egy osztályt képviselnek – a leírással keletkező osztálynak a tartalmi ismertetőjegyei). Nevezik őket mellérendelő információkereső nyelveknek (vagy „mellérendelő” osztályozási rendszereknek) is. Ilyen minden szemantikai nyelv (tehát a deskriptoros és a tárgyszavas nyelvek), és minden kötött szótárba nem foglalt, illetve nem szabványosított nyelv, valamint az enumeratív nyelvek egy – túlnyomórészt természetes nyelven alapuló – része (például az egysze-

rű, kapcsolatokat nem tartalmazó tárgyszórendszerek).

Az osztályozási rendszereket a gyakorlatban mellérendelő módon (szűkebb értelemben vett „információkereső nyelvként”, vagyis több osztály mellérendelésével) is használják, de ez korántsem olyan mérvű, mint az igazi mellérendelő információkereső nyelvek esetében.⁸ Az ETO elődjét, a Tizedes Osztályozást (TO) eredetileg nem abból a célból készítették, hogy vele mellérendelő (leíró) módon osztályozzanak. Idővel azonban kiderült, hogy olykor elkerülhetetlen a mellérendelő használat, és ennek érdekében – már a későbbi ETO létrehozói – bevezették a kettőspontos jelzet-összekapcsolást, a viszonyítást (pl. „réz építőanyag” esetében 669.35:691.73), amit aztán a TO-ban is használni kezdtek. A viszonyítást ma már többnyire nem jelölik, a jelzeteket kettőspont nélkül egyszerűen csak felveszik a dokumentumtételbe, már csak azért is, mert az automatizált könyvtári rendszereket (egyes kivételes alkalmazásoktól eltekintve) sem készítik fel jelenleg a jelzet-, deskriptor- és tárgyszóláncokon belüli szintaktikai (az adott kérdés szavai közötti) kapcsolatok, az ún. relátorok és szerepjelölők kezelésére. Az ETO-val szemben az SZJ-számok rendszerét viszont csak a szó szűk értelmében vett „osztályozó” módon használják, egy adott szolgáltatás csak egyetlen SZJ-számmal jellemezhető.

Azok a köznévi deskriptorok, tárgyszavak stb., amelyekkel a információkereső nyelvek esetében az osztályt „leírják”, önmagukban persze szintén osztályok, de ez nem ellentmondás. Egy dokumentumot például besorolhatunk a „könyvtár, dokumentumszolgáltatás, elektronikus könyvtár” deskriptor- vagy tárgyszólánccal „leírt” osztályba, és innen nézve a lánc három eleme a lánc egészével megnevezett osztály három ismertetőjegye. Ugyanakkor a „könyvtár”, a „dokumentumszolgáltatás” és az „elektronikus könyvtár” is osztályok, amelyek az adott információkereső nyelvben a dokumentumokat besoroló osztályok elemei lesznek. Noha az ETO-val is lehet mellérendelő módon osztályozni (két- vagy három ETO-jelzet kettőspontos vagy sima összekapcsolásával), ez nem az ETO lényegi jellemzője, ezzel szemben a – szűkebb értelemben vett – információkereső nyelvek lényegi jellemzője. Adott esetben a deskriptor- vagy tárgyszólánc állhat egyetlen elemből (például „könyvtár”), maga a deskriptor vagy tárgyszó ebben az esetben egyszerre ismerv és osztály, mivel a szóban forgó „leírt” osztály (a „könyvtár”) az alkalmazott esetben egyetlen ismertetőjegyből áll.

Az ilyen értelemben vett információkereső nyelvek szavai („osztályai”) abban különböznek az osztályozási rendszerek „szavaitól” (osztályaitól), hogy az utóbbiak eredendő rendeltetése az osztályszerű (besoroló) használat, az előbbieké viszont a leíró jellegű használat, melynek csak az eredményeképpen keletkezik osztály. Ez az eltérés ugyan nagyon hajszaifinomnak tűnik, de okozója annak, hogy a nemzetközi szabványokban [2, 14] és a szaknyelvben is – ha nem is mindenütt és következetesen, illetve vitatható módon – információkereső nyelvekről beszélnek, ha a deskriptoros vagy a tárgyszavas nyelvekről van szó, és osztályozási rendszerről, ha a TO-ról és az ETO-ról van szó. És nem elsősorban azért, mert az utóbbi kettő szerkezete hierarchikus, mert mint láttuk, a fenti, szűkebb értelemben vett információkereső nyelv is lehet hierarchikus.

Ez az „osztályozási rendszerek versus (’tulajdonképpen’) információkereső nyelvek” felosztás teljesen más szempontot képvisel, mint az, amit az 1. táblázatban adtunk meg. A táblázatban ugyanis egységbe igyekeztünk foglalni minden olyan nyelvet, amelyet információkeresésre használnak, tekintet nélkül a szaknyelvben élő – ha nem is mindenki által kizárólagosnak elfogadott – terminológiára, amelyet az előző bekezdésben tárgyaltunk. Annak érdekében, hogy ez utóbbi terminológiában az osztályozási rendszernek és információkereső nyelvnek legyen közös fölérendeltje, alakult ki a dokumentációs nyelv (documentary language) fogalma, melynek tehát fajtái az információkereső nyelv és az osztályozási rendszer. Elvben van még egy harmadik fajtája is, az indexelő vagy dokumentumleíró nyelv (amellyel a dokumentumokat tartalmilag leírják).⁹ A könyvtári rendszerekben az indexelő és az információkereső nyelvek lényegében egybeesnek. Például ugyanazzal a tezauszba foglalt deskriptoros nyelvvel indexelik (dolgozzák fel tartalmilag) a dokumentumokat, mint amellyel keresnek is.

Azért esnek csak lényegében egybe, mert vannak kivételek: a felhasználó például kereshet teljesen szabadon választott keresőszavakkal olyan tárolóban, amelynek dokumentumait mondjuk deskriptorokkal indexelték. Az első lépésben tehát ilyenkor nem azonos a kereséshez használt és az indexeléshez használt nyelv, de az automatizált könyvtári rendszerekben a második lépésben már a pontos kereséshez alkalmas indexelő nyelvi kifejezések mutatóját ajánlja fel a rendszer, és ettől kezdve a két nyelv egybeesik.

Újabb a web dokumentumai („erőforrásai”) között végzendő automatikus tartalmi feltárás („szemantikus web”) és a tudásreprezentáció területén alkalmazott, valójában egyáltalán nem új (a tezauszokban már kezdettől fogva alkalmazott) hierarchikus struktúrákat „ontológiáknak” nevezik. A bennük szereplő természetes nyelvű hierarchikusan rendezett kifejezések is információkereső nyelvi szavak.¹⁰

A szavak használat szempontjából vett rendezését a 2. (példában nem teljes körű) táblázat tartalmazza.

2. táblázat

A dokumentációs nyelvek szavai (felosztás a használat szempontjából)

dokumentációs nyelv szava
1. osztályozási rendszer jelzete
1.1 ETO-jelzet
1.2 SZJ-szám
1.3 NSZO jelzete
1.4 FEOR kódszava
2. 'tulajdonképpen' információkereső nyelv szava
2.1 elsődleges azonosító
2.1.1 országnév-kódszó
2.1.2 ISBN/ISSN
2.1.3 URL
2.1.4 postai irányítószám
2.1.5 telefonszám
2.1.6 személyi szám
2.1.7 adószám
2.1.8 adóazonosító jel
2.1.9 kötött szótár nélküli, szabványosított mesterséges nyelven alapuló szó
2.2 egységesített besorolási adat
2.3 deszkriptor/nemdeszkriptor
2.4 tárgyszó
2.5 szabadon választott keresőszó
3. dokumentumleíró nyelv szava

Kötött és szabad szavas keresés

A felhasznált nyelvek típusa alapján tehát egészen általános értelemben beszélhetünk szabványosított és nem szabványosított szavakkal végzett keresésről. Az előbbi nevezhetjük *kötött szavas keresésnek* (ide tartozik a jelzetek/kódszavak¹¹ szerinti keresés, a deszkriptoros és a tárgyszavas), továbbá az egységesített besorolási nevekkkel és címekkel (és könyvek esetében a könyv mutatószavai meg a tartalomjegyzék alapján) végzett keresés.

A kötött szavas keresés esetében különösen ki kell emelni a mutatók (indexek) szerepét és fontosságát. Annak érdekében ugyanis, hogy a felhasználó tudja is, milyen keresőszavak állnak rendelkezésére, fel kell kínálni a választékukat.

A sikeres keresés ugyanis az on-line mutatóban végzett böngészéssel kezdődik.¹² Enélkül legfeljebb véletlenül található ki, hogy az „Új Pedagógiai Szemle” cikkei egyes időszakokban adott adatbázisban az „Új Ped. Szle” formában kell keresni, a Tudományos és Műszaki Tájékoztatás című folyóiratot pedig a P 653 raktári jelzettel. Ráadásul a kötött szótárak szavai jelentős részben több szóból állnak, ezért mind a szavankénti, mind pedig a szóösszetételenkénti mutatókra egyformán nagy szükség van (pontosabban lenne – különös tekintettel a jelenlegi rendszerek e téren nyomorúságos kínálatára).

A tárgyszavakat vagy tezauszt használó rendszerek esetében pedig a mutató csak szükséges, de nem elégséges feltétel: ilyenkor át kell tudni térni a mutatóból a tárgyszavak vagy a tezausz strukturált állományára, hogy a kapcsolatok alapján is böngészni lehessen. Ami ezt illeti, a könyvtári rendszerek e tekintetben nem hogy tezauszokat nem tudnak még kezelni megfelelő módon, de sokszor még mutatókat sem kínálnak fel minden keresőszó-típus esetén.¹³

A keresés megnevezése a nem szabványosított szavakkal végzett keresés esetében sokszor hibás, mivel széles körben elterjedtek a „szabad tárgyszó” és a „szabad tárgyszavas keresés” megnevezések. Ez azért fából vaskarika, mert a tárgyszó mindig szabványosított információkereső nyelvi kifejezés.¹⁴ A „tárgyszavas keresés” tehát szabványosított információkereső nyelvi kifejezéssel végzett keresést jelent, a „szabad” jelző pedig azt jelenti, hogy nem szabványosított keresőszóval végzünk keresést. Következésképpen az említett megnevezés azt jelenti, hogy „nem szabványosított szabványosított kifejezéssel végzett keresés”, ami nonszensz.

A helyes megnevezés azon alapszik, hogy ebben az esetben olyan kifejezéssel keresünk, amelyet sem kötött szótárból nem választunk (például tezausztól, tárgyszórendszerrel függetlenül megadjuk a keresőképpen, kereső „mondatban”, hogy „kovács”, amin szakmát értünk, de e tényt sem jelöljük), sem olyan szabályok szerint nem képezzük a keresőszót, amelyek a kötött szótár nélküli,

de szabványosított információkereső nyelvre jellemzők (például megadjuk, hogy „kovács”, és azt is, hogy ezen családnevet értünk).

Az első esetben lehet, sőt minden bizonnyal létezik olyan kötött szótár (tárgyszójegyzék, teaurusz vagy például a „Foglalkozások Egységes Osztályozási Rendszere” [FEOR]), amely ezt a szót vagy a jelzetét tartalmazza, de ezek egyikének szava sincs hozzárendelve a dokumentumtétel ama mezőjéhez, melynek szavait az összehasonlításra kiválasztjuk (a keresés és a mező kapcsolatával a következő fejezetben foglalkozunk). Létezhetnek különféle „Kovács” rendszói (vagyis egységesített) részt tartalmazó, egységesített besorolási nevek is, de a kereséshez használt „kovács” szó maga nem egységesített alak.

Ilyenkor valójában teljesen szabadon választjuk meg a keresőszót (nem szótárból, és tekintet nélkül bármiféle szabályozásra), s a keresésnek ezt a módját „szabad szavas keresésnek” nevezzük. Az on-line katalógusokban (és más, kereshető adatbázisokban) ugyan még ezekkel a dokumentumokhoz hozzárendelt szabadon választott szavakkal is megjelenhet mutató, amelyet a könyvtári rendszer automatikusan hoz létre. Ez természetesen már afféle szótárnak is tekinthető, de e szótár kifejezéseire továbbra is érvényes, hogy szabadon választották őket, a velük végzett keresés tehát szabad szavas.

A korszerű on-line könyvtári rendszerekben gyakorlatilag minden ismérvből készülhet on-line mutató, tehát azokból a fajta ismérvekből is, amelyek szabadon választott szavakból keletkeztek. Ezért aztán maga a keresés a gyakorlatban már nem tökéletesen „szabad szavas”, legfeljebb a tartalmi feltárás lehet ilyen. Teljesen szabad szavas keresés valósul meg a weben az indexelő szolgáltatásokban (AltaVizsla, Ariadnet, Heuréka, Google): ott a kereső mindig teljesen szabadon választ keresőszót, noha vannak rendszerek, melyek például tezaurusszal is támogatják a keresőt.¹⁵

Az elterjedt hibás szóhasználat miatt tehát külön is kiemeljük:

A „szabad tárgyszó” és a „szabad tárgyszavas” kifejezések helyett a „szabadon választott szó” (esetleg röviden: „szabad szó”) és a „szabad szavas” kifejezések használandók.

A keresőszó fajtái szerinti kereséseket táblázatban is összefoglaltuk (3. táblázat).

3. táblázat

A keresések fajtái a keresőszavak alapján

kötött szavas keresés

jelzetek szerinti keresés (pl. ETO szerinti keresés, kódszavas keresés)
tárgyszavas keresés
deszkriptoros keresés
keresés egységesített besorolási nevekkkel vagy címeikkel¹⁶

szabad szavas keresés

A keresés fajtái a dokumentum szempontjából

Ahhoz, hogy egy dokumentum kereshető legyen, a dokumentumot leíró – a dokumentumok szövegét nem tartalmazó – dokumentumtételhez hozzá kell kapcsolni valamilyen információkereső nyelvi szót (vagy szavakat), melyek alapján kereshetők. A dokumentumtétel gyakran tartalmaznak rövidebb-hosszabb tartalmi kivonatokat (annotációkat, referátumokat), melyek szövegszavai alapján ugyancsak kereshető a tétel. A web dokumentumai és a digitalizált dokumentumok esetében¹⁷ a dokumentumtétel egyben a dokumentum teljes – természetes nyelvű – szövege vagy annak egy része, e szövegszavak alapján a keresés ugyancsak lehetséges. Mind a tartalmi kivonatokat, mind a teljes dokumentumszöveg esetében a dokumentum nyelve és az információkereső nyelv azonosága csak korlátozott, hiszen a kereső a szabadon vagy szótárból választott szavairól nem tudhatja előre, hogy ezek egyáltalán előfordulnak a tartalmi kivonatban vagy a dokumentumban. Más a helyzet, ha a tartalmi kivonat vagy a dokumentumszöveg szavait automatizáltan indexelik, és mutatót kínálnak fel. A web HTML dokumentumai esetében az indexelő szolgáltatások az indexelést elvégzik, de a mai szoftvertechnikai körülmények között az így indexelt állomány mutatóként való felkínálása még nem valósítható meg a keresési sebesség jelentős csökkenése nélkül, ezért a kereső mintegy „vaktában” adja meg webes kereséskor szabadon választott keresőszavait.

A tételhez kapcsolandó szavakat, mint láttuk, választhatjuk kötött szótárból vagy választhatjuk szabadon. A dokumentumok felől nézve – pontos

fogalmazás esetén (jóllehet adott esetben ugyanazon nyelvről van szó) nem információkereső nyelvekről, hanem dokumentumleíró nyelvekről beszélhetünk (a kifejezést az információkereső nyelv szavaival foglalkozó első fejezetben már tárgyaltuk). A sikeres keresés triviális előfeltétele, hogy a két nyelvnek adott információkeresés alkalmából azonosnak kell lennie. Más szóval a dokumentumot elvileg ugyanazon a nyelven kell keresni, mint amilyen nyelven a dokumentumtételben kereshető szavak (pl. szövegszavak) vannak, vagyis mint amilyen nyelven osztályozták/indexelték (= tartalmilag leírták). Ez az előző fejezet alapján lehet kötött szótárba foglalt vagy szótárba nem foglalt, szabványosított vagy nem szabványosított nyelv. Tartalmi kivonat vagy teljes dokumentumszöveg esetében a keresőrendszerek egy részében meghatározható, hogy a szabadon választott szavakkal a dokumentumtételnek abban a részében akarunk-e keresni, amely a kivonatot vagy a teljes szöveget tartalmazza, vagy abban a részében (ha ilyen van), amelybe a tartalmi feltárásakor dokumentumleíró nyelven szavakat vittünk be.

Mindegyik esetben arról van szó, hogy a dokumentum leírását tartalmazó dokumentumtétel mezőből épül fel, pl. a főcím mezője, a tartalmi kivonat mezője, az ETO-jelzet mezője, a tárgyi kifejezések mezője, a teljes dokumentumszöveg mezője. (Olyan dokumentumtételek esetében, amelyek a teljes szöveget tartalmazzák, előfordulhat, hogy a tétel csak a teljes szöveget tartalmazza; ebben az esetben egyetlen mezős tételről beszélhetünk.) A mezők lehetnek kereshetők vagy nem, azaz a mezők szavai lehetnek ismérvek vagy nem. A tartalmi kivonat vagy a teljes dokumentumszöveg esetében (ha mezőjük kereshető) az ismérvek maguk a szövegszavak. Egy dokumentumot többfajta információkereső nyelvvel is leírhatunk, de akkor annyi mezőt (pl. ETO-jelzet mezőt, tárgyszómezőt, szabadon választott hozzá kapcsolt szavak mezőjét, szövegszavas mezőt) kell ehhez meghatározni, ahány különféle nyelvet használunk. Ezeknek a mezőknek (adatelemeknek) a keresésbe bevonható egységeit (kezdvé a jelzektől a tárgyszavakon és deszkriptorokon át a szövegszavakig) nevezzük a dokumentum ismérveinek.¹⁸

A fentiek értelmében bizonyos mezőkben nincsenek előre megadott ismérvek, mégis bevonhatók a keresésbe. Az ilyen mezők tartalma valódi vagy annak tekintett folytonos szöveg (illetve szövegrész), melynek bármelyik szavával vagy össze-

tett kifejezésével szabadon keresni lehet. Ezek a szövegszavak. A szabad szavas keresés könyvtári és hasonló viszonyok között részben, és – ez későbbi fejlemény – a webes indexelő szolgáltatásokban kizárólag ezeken a mezőkön, tehát a szövegszavakon alapszik. Ilyen mezők elsősorban a címetek tartalmazó mezők (a főcím, alcím, párhuzamos cím adatelemei, illetve a HTML dokumentumok fejrészében szereplő <cím> mező) és a tartalmi kivonatok mezői, illetve a HTML dokumentum szövege. A szabad szavas keresés könyvtári rendszerekben egyrészt a címben (esetleg a tartalmi kivonatban) szereplő tetszés szerinti szavakon alapszik, másrészt azokon a – kifejezetten a keresés céljait szolgáló – szavakon, amelyeket önálló mezőbe szabadon választva vittek be. A rendszerek használói, de maguk a készítők is főleg ezeket az utóbbi szavakat nevezik tévesen „szabad tárgyszavaknak,” s mivel szinte mindenütt kihasználják ezt a lehetőséget, a hibás szóhasználat is rendkívül elterjedt.

A tárgyszavakat, deszkriptorokat vagy szabadon választott szavakat tartalmazó ismérvmezők is kezelhetők adott rendszerben a kereséskor szöveges mezőként; ilyenkor az ismérv maga számít folytonos szövegnek, és egyes elkülönülő elemei szövegszónak. (Például egy deszkriptormezőben szereplő „adásvételi szerződés” deszkriptor esetén, ha szöveges mezőként kezeli a kereséskor a rendszer, az adott dokumentum [tétel] a „vétel” vagy a „szerződés” kifejezésekre is releváns lehet.)

Igazán innovatív lehetőség volna, ha az ETO-jelzetekkel osztályozott dokumentumtételeket a hozzájuk kapcsolt jelzetek szöveges magyarázataiban szereplő szövegszavak alapján is lehetne szabad szavas kereséssel keresni. Ez számítástechnikailag egyáltalán nem ördögösség, feltehetően azért hiányzik mégis ez a valójában pompás lehetőség, mert senkinek nem jutott az eszébe, és mert a technikai és elvi újdonságok abszolutizálása vagy félremagyarázása miatt „lekezelik” az ETO-t és szöveges megfelelőjét.

Mindazokat a fent leírt (a) és (b) mezőket, amelyek a keresésre felhasználhatók, nevezzük némi elnagyoltsággal keres(het)ő mezőknek. Valójában nem ezeket a mezőket keressük (ahogy ezt a rövid megnevezés sugallja), hanem a keresés, s ezen belül az összehasonlítás ezeknek a mezőknek az ismérvein/tartalmán alapul. A pontos megnevezés „kereséshez használható mező” volna.

Az első esetben (a) *ismérvmezős keresés*, a második esetben (b) *szövegmezős keresés* játszódik le. Pontosabban: az első esetben olyan ismérvről van szó, amelyet a kereshetőség érdekében „bevittek”, a második esetben viszont ismérvvé válhattott olyan szó, melyet eredetileg nem ebből a célból vittek be. A szövegmezős keresés egyik leggyakoribb változata az ún. *címszavas keresés*, amelynek esetében a szövegmező a cím. Ez nem azonos a *cím szerinti kereséssel*, amelyet elvileg a teljes cím, mint besorolási adat alapján végeznek, és lényegében kötött szavas keresés, mert a teljes, szükség szerint egységesített címre vonatkozik. A címszavas keresés gyakorlatilag mindig szabad szavas keresés a címen belüli szavak vagy együttesük bármelyikével. A kétféle keresés a gyakorlatban szorosan kapcsolódhat egymáshoz, mivel a címszavas kereséssel – korszerű on-line katalógusokban – találhatjuk meg az egységesített címek választékát (feltéve, ha a könyvtárban egységesített besorolási adatokat egyáltalán használnak), melyből a keresett címet kiválasztva hajtjuk végre a cím szerinti keresést. A helyzet értelem-szerűen ugyanez a személyek és testületek egységesített nevei esetében is.

A keresés összetett fajtái

A valóságban a keresés általában vagy legalábbis igen gyakran se nem pusztán kötött szavas és szabad szavas, se nem pusztán ismérvmezős és szövegmezős, hanem ezek kombinációja. Vagyis kereshetünk kötött szavakkal ismérvmezőben és szövegmezőben, és kereshetünk szabadon választott keresőszavakkal ismérvmezőben és szövegmezőben egyaránt. A szabadon választott szavakkal végzett keresés irányulhat mind a dokumentum kifejezetten ismérveket tartalmazó mezőinek tartalmára, mind pedig szöveges mezőinek tartalmára. Kötött szótárak használata esetében a keresés célszerű módon általában csak az ismérvmezőkre irányul (noha elvileg irányulhatna a szöveges mezőkre is), a szabadon választott szavak esetében irányulhat az ismérvmezőkre vagy a szöveges mezőkre, de általában inkább az utóbbiakra. Azaz lehet *kötött szavas ismérvmezős* (és *kötött szavas szövegmezős*) keresés, továbbá *szabad szavas ismérvmezős*, és főleg *szabad szavas szövegmezős keresés*.

A 4. táblázatban láthatók a fontosabb esetek, melyek közül a gyakorlatban leginkább előfordulókat félkövéren írtuk.

4. táblázat

A keresések fajtái

	ismérvmezős keresés	szövegme- zős keresés
kötött szavas keresés	i	i
jelzet szerinti keresés	i	-
deszkriptoros, tárgyszavas k.	i	i
cím szerinti keresés	-	i
keresés (internet)	-	i
szabad szavas keresés	i	i
címszavas keresés	i	i
keresés (internet)	i	i

Kitüntetett szavak

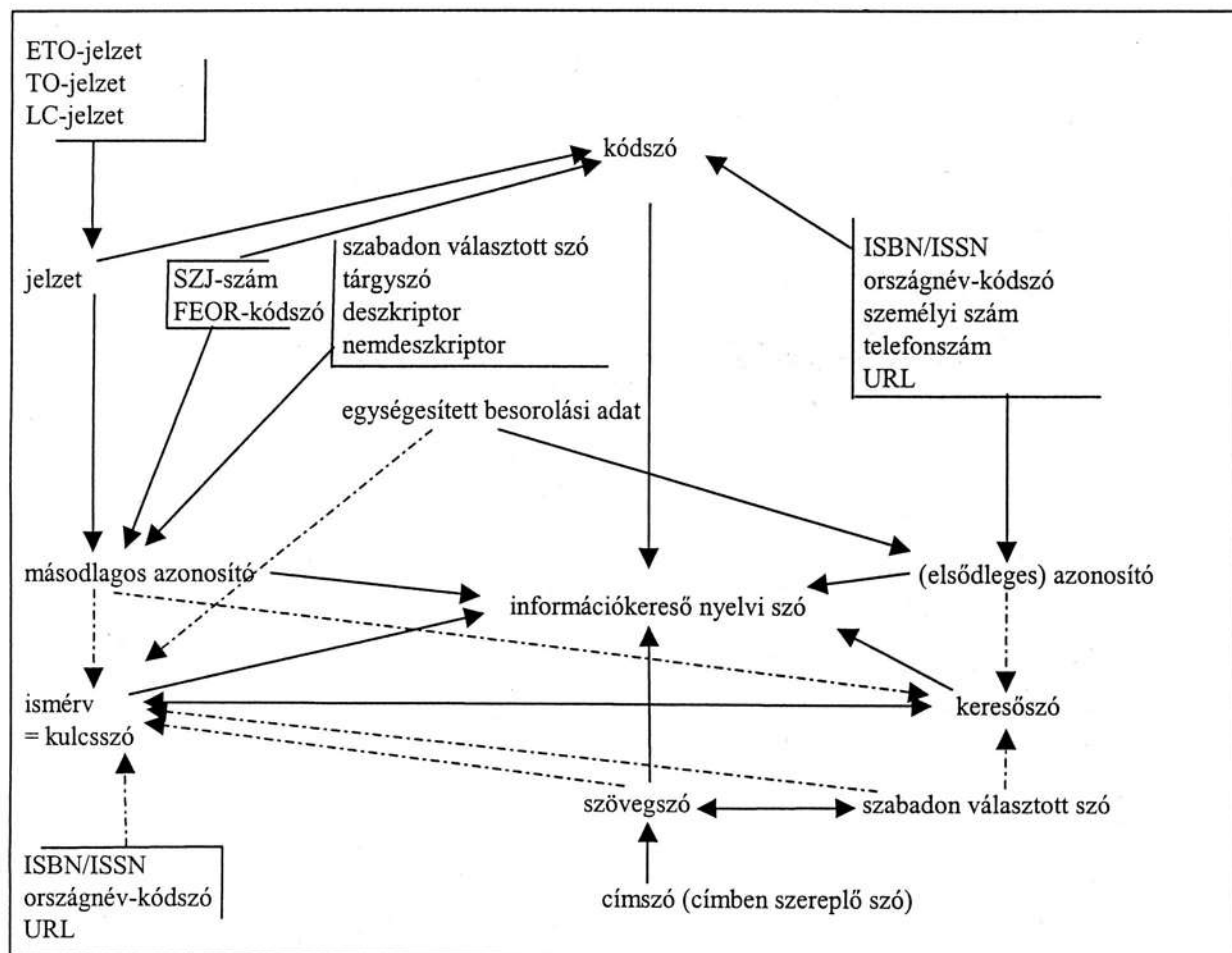
Az eddigiekben számos olyan szó merült fel, melynek használata nem kerülhető el, ha az információkereséssel foglalkozunk. Ezekről szólunk az alábbiakban. Ilyen először is maga az információkereső nyelvi szó, melynek egyik alkalmazása a keresőszó. Hozzájuk társulnak a speciális információkereső nyelvi szavak, mint a jelzet, a tárgyszó, a deszkriptor és a szabadon választott szó („szabad szó”).

Külön csoportot képviselnek a dokumentummal összefüggő információkereső nyelvi szavak, mint az ismérv és a szövegmező. A szövegmező egyik kitüntetett fajtája a címszó. Az ismérvnek megfelelő és a számítástechnikai gyakorlatban használt kifejezés a kulcsszó. Abból a megfontolásból született, hogy az ismérvek hozzáférési elemek, afféle „kulcsok” a dokumentumtételhez, az ismérv a kereséskor a „nyitja” a dokumentumtételnek.¹⁹ Ismérv és kulcsszó tehát szinonimák: az elsőt a könyvtári információkeresés és osztályozás gyakorlatában, a másodikat pedig főleg a web dokumentumaiban (példa a HTML dokumentumok <kulcsszó> mezője), a számítástechnikában és a rendszerelméletben használják.

Ismérvek/kulcsszavak szerepét játszhatják a *másodlagos azonosítók* (az ETO-jelzet, a tárgyszó, a deszkriptor stb.) és az elsődleges azonosítók ama fajtái, melyek a könyvtári és webes gyakorlatban előfordulnak (egységesített besorolási adatok, ISBN/ISSN, országnév-kódszó, URL). A másodlagos azonosítókkal nem egyetlen egyed, hanem egyedek osztálya azonosítható, szemben az *elsőd-*

leges azonosítóval, amely mindig csak egyetlen és csakis egyetlen egyedet azonosít. A gyakorlatban azonosítón mindig elsődleges azonosítót (elsődleges ismérvet, elsődleges kulcsot) értenek. Az elsődleges és másodlagos azonosító használatának egyrészt a szoftverkészítéskor (a jól megválasztott azonosítóknak fontos szerepük van a programok hatékony működésében és karbantartásában), más-

részt az elméleti irodalomban van szerepe. Ismérv/kulcsszó nemcsak dokumentumtételhez kapcsolt (elsődleges) azonosító, jelzet, tárgyszó, deszkriptor vagy szabadon választott szó lehet, hanem a tétel bármely szövegszava. Különösen gyakori, hogy a címszó is az ismérv/kulcsszó szerepét játssza. A kitüntetett szavakat címkéztük, irányított gráf formában foglaltuk össze (1. ábra).



1. ábra Kitüntetett információkereső nyelvi szavak

A gráf olvasata:

- az ETO-jelzet, a TO-jelzet és az LC-jelzet jelzetek;
- a jelzet, az országnév-kódzó, az SZJ-szám, a FEOR-kódzó, az ISBN/ISSN, a személyi szám, a telefonszám és az URL kódzavak;
- a jelzet, az SZJ-szám, a FEOR-kódzó, a tárgyszó, deszkriptor és nemdeszkriptor másodlagos azonosító;
- az egységesített besorolási adat, az ISBN/ISSN, a személyi szám, a telefonszám és az URL elsődleges azonosító;
- az elsődleges és a másodlagos azonosító, az ismérv és a keresőszó, a kódzó és a szövegszó információkereső nyelvi szó (a kódzó fajtái ugyanis lehetnek el-

- sődleges és másodlagos azonosítók, ezért maga a kódzó e kettő közös fölérendeltje lehet csupán);
- a címszó (címben szereplő szó) szövegszó;
- a szabadon választott szó játszhatja a keresőszó (ezzel végezhető a szabad szavas keresés) és az ismérv szerepét, és fajtája a másodlagos azonosítónak;
- a szabadon választott szó egyben tekinthető a szövegszó afféle keresőszavi megfelelőjének, melyet az „egyéb rokonsága” relációval jelöltünk;
- az elsődleges és a másodlagos azonosító, továbbá a szabadon választott szó játszhatja a keresőszó szerepét, a másodlagos azonosító, az egységesített besorolási adat, a szövegszó, az ISBN/ISSN, az ország-

név-kódszó és az URL játszhatják az ismerv szerepét (az általunk tárgyalt keretek között nem minden elsődleges azonosítót használnak dokumentumok ismervként, ezért nem lehetett az elsődleges azonosítót, hanem csak meghatározott fajtát a „szerepe” relációval az ismervhez kapcsolni – a telefonszám vagy a személyi szám nem válhat dokumentum ismervvé);

- az ismerv szerepét játszhatja egységesített besorolási adat, másodlagos azonosító (természetes nyelvű kifejezés) és olyan kódszó, amely dokumentumok tartalmát vagy formai tulajdonságait jellemzi, de az ismérvek maguk nem keresőszavak (mintegy az ellentétének tekinthető, amit az „egyéb rokonsága” relációval fejeztünk itt ki);
- a keresőszó szerepét játszhatja minden elsődleges és másodlagos azonosító.

A tanulmányban előforduló információkereső nyelvi szavak

adóazonosító jel (magánszemély adószáma): mesterséges nyelven alapuló, szabványosított, kötött szótáras, elsődleges azonosító [pl. 8255171954]

adószám (cég adószáma): mesterséges nyelven alapuló, szabványosított, kötött szótár nélküli, elsődleges azonosító [pl. 28341519–2–41]

országnevkód (pontosabban: országnevkódszó): mesterséges nyelven alapuló, szabványosított, kötött szótáras, enumeratív, elsődleges azonosító [pl. hu]

egységesített besorolási adat: természetes nyelven alapuló, szabványosított, kötött szótár nélküli, elsődleges azonosító [pl. Möriz Zsigmond (1879–1942)]

ETO-jelzet: mesterséges nyelven alapuló, szabványosított, kötött szótáras, hierarchikus, másodlagos azonosító, osztályozási rendszer „szava” [pl. 930.85(439)]

FEOR-szám (Foglalkozások Egységes Osztályozási Rendszere kódszáva): mesterséges nyelven alapuló, szabványosított, kötött szótáras, hierarchikus, másodlagos azonosító, osztályozási rendszer „szava” [pl. 7426]

ISBN/ISSN: mesterséges nyelven alapuló, szabványosított, kötött szótáras, elsődleges azonosító [pl. 963–571–418–1]

Kongresszusi Könyvtár osztályozási rendszerének jelzete (LC-jelzet): mesterséges nyelven alapuló, szabványosított, kötött szótáras, hierarchikus (eredetileg enumeratív), másodlagos azonosító, osztályozási rendszer „szava” [pl. TK–1602]

NSZO (Nemzetközi Szabadalmi Osztályozás) jelzete: mesterséges nyelven alapuló, szabványosított, kötött szótáras, hierarchikus, másodlagos azonosító, osztályozási rendszer „szava” [pl. A61k]

postai irányítószám: mesterséges nyelven alapuló, szabványosított, kötött szótáras, enumeratív (részben hierarchikus), elsődleges azonosító [pl. H 1827]

szabadszó („szabad tárgyszó”): természetes nyelven alapuló, nem szabványosított, kötött szótár nélküli, másodlagos azonosító [pl. „kovács”, „Kovács”]

személyi szám: mesterséges nyelven alapuló, szabványosított, kötött szótár nélküli, elsődleges azonosító [pl. 1 590123 0097]

SZJ-szám (Szolgáltatások Jegyzékének száma): mesterséges nyelven alapuló, szabványosított, kötött szótáras, hierarchikus, másodlagos azonosító, osztályozási rendszer „szava” [pl. 92.51.11.0]

szövegszó: természetes nyelven alapuló, nem szabványosított, kötött szótár nélküli, másodlagos azonosító [pl. „kovács”, vagy „Kovács”]

tárgymutatószó könyvben: természetes nyelven alapuló, nem szabványosított, kötött szótáras, enumeratív, másodlagos azonosító [pl. kovács]

tárgyszó kapcsolatok nélkül: természetes nyelven alapuló, szemantikai, szabványosított, kötött szótáras, enumeratív, másodlagos azonosító [pl. kovács]

tárgyszó kapcsolatokkal: természetes nyelven alapuló, szemantikai, szabványosított, kötött szótáras, szemantikai, másodlagos azonosító [pl. kovács]

tartalomjegyzék eleme: természetes nyelven alapuló, nem szabványosított, kötött szótáras, általában hierarchikus, elsődleges azonosító [pl. „A tanulmányban előforduló információkereső nyelvi szavak”]

telefonszám: mesterséges nyelven alapuló, szabványosított, kötött szótáras, enumeratív, elsődleges azonosító [pl. 36 1 224–3738]

tezaurusz lexikai egysége (deszkriptor és nem-deszkriptor): természetes nyelven alapuló, szabványosított, kötött szótáras, szemantikai, másodlagos azonosító [pl. kovács]

TO-jelzet: mesterséges nyelven alapuló, szabványosított, kötött szótáras, hierarchikus, másodlagos azonosító, osztályozási rendszer „szava” [pl. 621.200.15]

URL: mesterséges nyelven alapuló, szabványosított, kötött szótár nélküli, elsődleges azonosító [pl. <http://www.net.hu/search/>]

Jegyzetek

- ¹ Információkeresésen e tanulmányban dokumentumok hivatkozásainak (a dokumentumok bibliográfiai szabványok vagy egyéb szabályok szerinti szűkebb vagy bővebb leírásainak, az ún. dokumentumtétteleknek) a keresését értjük, és a továbbiakban csak a keresés kifejezést használjuk. Könyvtári rendszerekben elsősorban ilyen kereséseket végeznek. A szakirodalomban nevezik hivatkozáskeresésnek (reference retrieval, Dokumentnachweisretrieval). Az információkeresésen belül megkülönböztetik még az elsődleges dokumentumok keresését (ezt nevezik dokumentumkeresésnek, document retrieval) és a faktografikus vagy adatkeresést (data retrieval, fact retrieval). Részletesen lásd [13, p. 14.]. Magának az általános értelemben vett keresésnek az információkeresés mellett még számos más szempontból nézve is vannak fajtái. Például a keresési folyamat alapján bináris, láncolt, lépésenkénti keresés stb., melyekkel itt nem foglalkozunk. Terminológiájukat lásd [14].

- ² A természetes és mesterséges nyelv meghatározását lásd [1].
- ³ Az összehasonítás, s vele a kereső- és dokumentumkép szerepét illetően lásd [13, p. 337–338].
- ⁴ A szóösszetételekből álló keresőszavakra még a „Kötött és szabad szavas keresés” című fejezetben kitérünk.
- ⁵ A deskriptor (és nemdeszkriptor) kifejezés elsősorban a tezauszok két fajta lexikai egységét jelenti, de általánosabb értelmű jelentése is van. Ez utóbbi esetben minden szabványosított információkereső nyelv előírt kifejezését jelenti [14, p. 113].
- ⁶ Az „utalás”, „utaló” valójában nemcsak a „lásd”, olykor a „lásd még” kapcsolat megnevezése, hanem a katalógusban (kereső eszközben) szereplő teljes utalótétel neve is. Az utalótétel (valójában információkereső nyelvi szócikk) ugyanis nemcsak azt tartalmazza, hogy „lásd” vagy „lásd még”, hanem azt is, hogy ez a reláció milyen szóról milyen szóra mutat. Mindezt a hagyományos cédulakatalógusokban külön katalóguscédula tartalmazta. A cédulakatalógusok háttérbe szorulásával az „utalás” és az „utaló” kifejezések is eltűnően vannak, mivel on-line katalógusok mutatóiban a különféle utalótételek nem jelennek meg olyan elkülönült formában, mint egy cédulakatalógusban.
- ⁷ Az igazgatási gyakorlatban a hierarchikus rendszereket regiszternek, ritkábban nomenklatúrának szokták nevezni (a FEOR mellett ilyen pl. a TEÁOR [Tevékenységek Egységes Ágazati Osztályozási Rendszere], a BTO [Belföldi Termékosztályozás], a „Cégregiszter” [a cégeket nyilvántartó rendszer]).
- ⁸ Az osztályozási rendszer vs. információkereső nyelv dichotómiából következik, hogy mindkettőre mindig szükség lesz a tartalmi feltárás és keresés területén. Ezért az ETO időszerűsége nem fog csökkenni, ellenkezőleg: gyakorlatilag az egyetlen, nélkülözhetetlen egyetemes könyvtári osztályozási rendszer, és valószínűleg most már hosszú időre az is marad [10].
- ⁹ Ez az – eleve nem egyértelmű, és az 1. táblázatban tárgyalt „információkereső nyelvi” gyűjtőfogalomtól nehezen vagy sehogy sem elhatárolható – fogalom az 1970-es években született, és hamar szabványosították. Ez az a nyelv, amelyet adott rendszerben az információk leírására, tárolására és keresésére használnak, lásd [2] és [14]. Az is tény, hogy szemben a 20. század 70-es és 80-as éveivel, ma a „dokumentációs nyelv” kifejezése – mint maga a „dokumentáció” fogalma is – kikopott a szakmai használatból, s vele az osztályozási rendszer és információkereső nyelv megkülönböztetésének a jelentősége is, mivel az on-line könyvtári (és más) rendszerekben eleve a mellérendelő használat a döntő.
- ¹⁰ Az ilyen ontológiákhoz sokszor formátumleíró nyelv is tartozik. Azonban nemcsak az ontológiát képviselő hierarchikus szótárt és számítógépes formátumának együttes rendszerét nevezik „ontológiának”, hanem a webkatalógusok természetes nyelvű osztályozási rendszereit is, mint amilyen például a Yahoo! vagy a magyar AltaVizsla „ontológiája” (sőt, felületesen minden generikus és egyéb hierarchiát is). Eszerint az ál-

talánosabb „ontológia fogalom” szerint azonban maga az ETO is „ontológia”. Az ontológiák kérdését összefoglalóan John F. Sowa tárgyalja [6, p. 51–123.].

- ¹¹ A kódszó helyett – hibásan – a „kód” kifejezést szokták használni. Ez azért helytelen, mert az utóbbi kifejezés nem az egyes kódszavakat, hanem az egész – többek között a kódszavakból, azok magyarázatából, a kódszavak képzési szabályaiból álló – rendszert jelenti. Innen nézve tehát a „kódrendszer” tautológia, hiszen magának a kódnak „mesterséges nyelven alapuló, meghatározott szabályok szerint képzett nyelvi rendszer” a jelentése. A „kód” matematikai-kibernetikai jelentése még ennél is szigorúbb („operátor, illetve szabályrendszer, amelynek az a rendeltetése, hogy adatokat karakterek halmazából vett elemsorozatok, szavak segítségével ábrázoljon”) [1, p. 60.]. A könyvtári-információs gyakorlatban alkalmazott kódszavakra a jelzet kifejezést használják (tehát nem minden kódszó jelzet, de minden jelzet kódszó). Innen nézve például az ETO egésze kód, és az ETO-számok (jelzetek) kódszavak.
- ¹² Ahogy erre Jacsó Péter cikkének címe is utal [3] (a kérdést átfogóbban is tárgyalja Jacsó [4, p. 199–218.]). Az on-line mutatók alapján végzett keresések gyakorlati tapasztalataiból ad gazdag ízelítőt Murányi Péter összehasonlító tanulmánya [5].
- ¹³ A tezauszok és tárgyszavas rendszerek felhasználói felületeivel részletesen foglalkozott [12].
- ¹⁴ A tárgyszóról részletesen lásd [11, p. 44–48.].
- ¹⁵ A példaként felsorolt nevek a szolgáltató rendszer nevei. E rendszereken belül működnek az indexelő programok, az ún. „keresőmotorok”, melyeknek általában nincs neve, ezért ezeket is a szolgáltatás néven tartják nyilván, vagy azt mondják, hogy pl. az „Ariadnet keresője”. Részletesebben lásd [8] és [9].
- ¹⁶ Az automatizált könyvtári rendszerekben lassan terjed az a lehetőség, hogy a kereséshez ne csak a deskriptorokat, hanem a nemdeszkriptorokat is föl lehessen használni (ilyenkor a rendszer maga tér át automatikusan az általa kezelt tezausz alapján a deskriptorra, vagy megadható, hogy csak azok a találatok jelenjenek meg, amelyeket a nemdeszkriptorokkal osztályoztak). Értelemszerűen ugyanaz vonatkozik a besorolási nevekre és címekre („utalókra”), amelyek az egységesített formákra utalnak [13, p. 139.].
- ¹⁷ Szemben a hagyományos könyvtári dokumentumokkal (nyomtatott és kéziratos könyvekkel, folyóiratokkal, térképekkel, kottákkal stb.), valamint a hagyományos vizuális és auditív dokumentumokkal (fényképek, mozgóképek, hangjátékok stb.) az elektronikus/digitális dokumentumok, webdokumentumok esetében az „erőforrás”, a „forrás” kifejezés használatát szorgalmazzák a nem könyvtári szakterületeken, például a számítástechnikában. Ez mindaddig nem baj (csak felesleges), ameddig a dokumentum, illetve a „forrás” kifejezések használói kölcsönösen tisztában vannak azzal, hogy a nevezett két fogalom terjedelme azonos.

- ¹⁸ Részletesen tárgyalja [13, p. 16. és 38.]. A mezők és adatelemek összefüggését illetően lásd [7].
- ¹⁹ Meg kell említeni, hogy a „kulcsszó” még ezen (vagyis a „számítástechnikai változat értelmén”) kívül is – sajnos – rengeteg értelemben használatos, így pl. a szabványosított természetes nyelvű információkeresés „tárgyszójelöltjeként”, szövegszó értelemben stb.

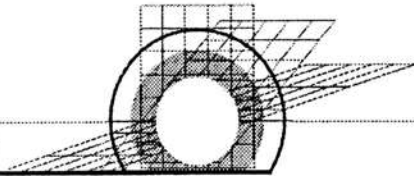
Irodalom

- [1] FREY Tamás–SZELEZSÁN János: Matematikai kibernetika. Budapest, Akadémiai K., 1973. (Műszaki értelmező szótár 34.) 120 p.
- [2] ISO 5127/6 – 1983. Documentation and information – Vocabulary – Part 6: Documentary languages.
- [3] JACSÓ Péter: Savvy searching starts with browsing. = Online & CD-ROM Review, 23. köt. 3. sz. 1999. p. 169–172.
- [4] JACSÓ Péter: Content evaluation of textual CD-ROM and web databases. Englewood, Colo., Libraries Unlimited, 2001. 276 p.
- [5] MURÁNYI Péter: Az időszaki kiadványok cikkeinek feldolgozása a magyar és a lett nemzeti bibliográfiai adatbázisokban. = Könyvtári Figyelő, 48. köt. 3. sz. 2002. p. 511–525.
- [6] SOWA, John F.: Knowledge representation. Logical, philosophical and computational foundations. Pacific Grove, Brooks/Cole, 2000. 594 p.

- [7] UNGVÁRY Rudolf: Dokumentációs és könyvtári szervezeti rendszerek elemzése. = Tudományos és Műszaki Tájékoztatás, 46. köt. 9–10. sz. 1999. p. 355–376.
- [8] UNGVÁRY Rudolf: A tartalom szerinti információkeresés az interneten. I. Indexelő szolgáltatások. = Tudományos és Műszaki Tájékoztatás, 47. köt. 1. sz. 2000. p. 3–19.
- [9] UNGVÁRY Rudolf: A tartalom szerinti információkeresés az interneten. II. Internetkatalógusok. = Tudományos és Műszaki Tájékoztatás, 47. köt. 2. sz. 2000. p. 55–67.
- [10] UNGVÁRY Rudolf: Az ETO szükségessége. = Iskolakultúra, 10. köt. 4. sz. 2000. p. 27–31.
- [11] UNGVÁRY Rudolf–ORBÁN Éva: Osztályozás és információkeresés. Kommentált szöveggyűjtemény. Budapest, OSZK, 2001. 1. köt. Az osztályozás elmélete. 544 p.
- [12] UNGVÁRY Rudolf: A tezaurusz a felhasználói felületen. Az optimális tezauruszmegjelenítés problémái. = Tudományos és Műszaki Tájékoztatás, 48. köt. 3. sz. 2001. p. 99–108.
- [13] UNGVÁRY Rudolf–VAJDA Erik: Könyvtári információkeresés. 2. jav. kiad. Budapest, Typotex, 2002. 170 p.
- [14] WERSIG, Gernot–NEVELING, Ulrich: Terminology of documentation. Paris, The Unesco Press, 1976.

Beérkezett: 2003. IX. 9-én.

Rendezvénytár



Content Management Europe 2003, konferencia és kiállítás

London, 2003. december 2–4.
Szervező: Vernon Tolson
Business Development Manager
Tel.: +44 1932 730735
E-mail: Vtolson@imark.co.uk
URL: <http://www.cme-expo.co.uk>

Könyvtár és információ a multikulturális társadalmakban. BOBCATSSS 2004

Riga, 2004. január 26–28.
Szervezők: University of Latvia
és University College of Boras
E-mail: info@bobcatsss.com
URL: <http://www.bobcatsss.com>

Digitális könyvtárak. Nemzetközi konferencia (International conference on digital libraries: knowledge creation, preservation, access and management)

Újdelhi, 2004. február 24–27.
Szervező: ICDL 2004 Secretariat, TERI,
Darbari Seth Block, Habitat Place, Lodhi Road,
New Delhi – 110 003
India
Fax: +99 11 24682133
URL: <http://www.teriin.org/events/icdl>

Átalakulások – fejlődésben a könyvtár. Konferencia

Aarhus (Dánia), 2004. június 13–15.
Szervező: Rolf Hapel, Library Director
Aarhus Public Libraries
Tel.: +45 8940 9300
E-mail: hapel@bib.aarhus.dk
URL: www.aakb.dk/transformation