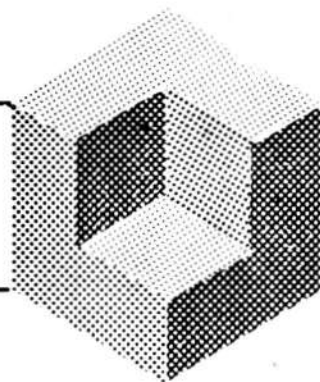


Követelmények és néhány megoldás



A DOKUMENTÁCIÓS CÉLÚ ADATBÁZIS-KEZELŐ RENDSZER

A rendszer helye, elvi működése és felépítése

Ungváry Rudolf

Országos Széchényi Könyvtár

1. A dokumentációs célú adatbázis-kezelő rendszer helye a programrendszerek között

1.1 A három programfajta

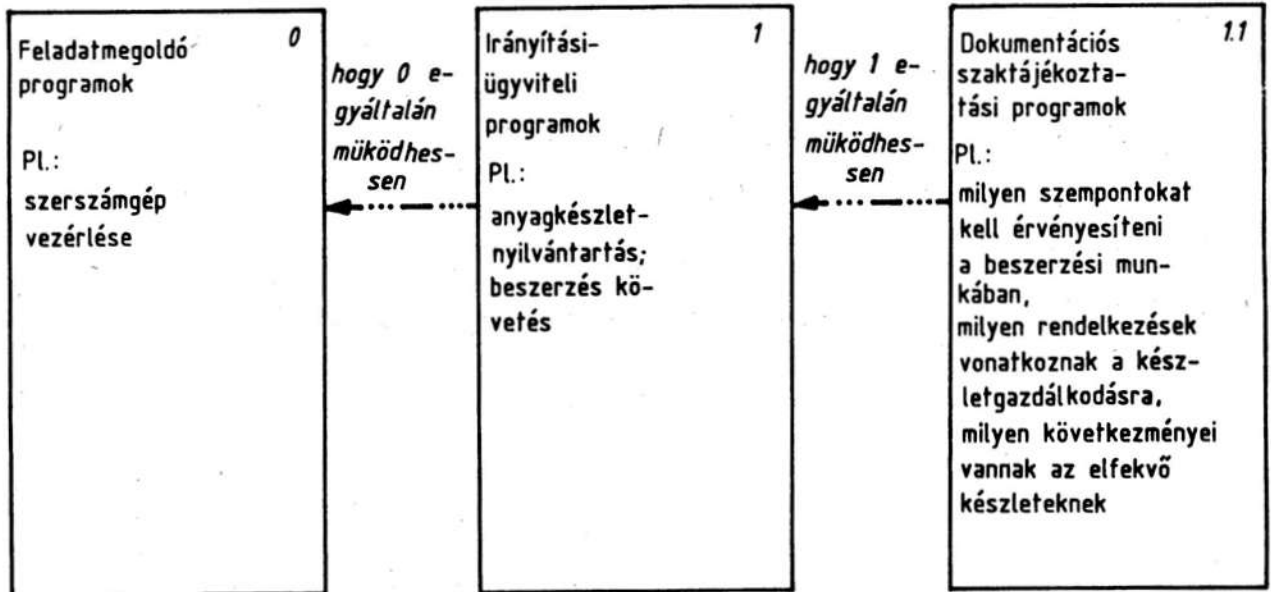
A programok feladata az, hogy valamit önállóan megoldjanak, valamilyen folyamatot, eljárást végrehajtsanak. A matematikai, szimulációs és a különféle – például termelési, élettani – folyamatokat irányító programokra jellemző, hogy az adatokat, amelyek az általuk kezelt rendszerbe bekerülnek, valamilyen másik adattá alakítják át, vagy a bekerülő adatok alapján előre meghatározott belső minta szerint kimenő (irányító) adatokat hoznak létre. Lényegük éppen ez az átalakítás, illetve a minta alapján előállított eredmény.

Vannak programok, melyek feladata elsősorban az, hogy adatokat tároljanak, mégpedig azért, hogy velük valaminek a programon kívüli, szellemi megoldását – például gazdasági, kulturális szervezetek irányítását és gazdálkodását – támogassák. Ilyenek a pénzügyi, munkaügyi, személyzeti stb. irányítás, a számvitel és a gazdálkodás, valamint a különféle nyilvántartások (állóeszköz, raktári, szállodai szoba, utas stb.) programjai. Amit ezek a programok megoldanak, az lényegében nem több, mint hogy a tárolóból a nyilvántartott adatokat a szellemi irányításhoz (megoldáshoz) szükséges formában és összefüggésekben kiadják. Éppen ezért e programok – különösen ami az adatok közötti kapcsolatokat és megjelenítésük módját illeti – elsősorban e rendszerek, azaz a gazdasági szervezetek igényeihez alkalmazkodtak. Az adatok közötti kapcsolatokat az adott szervezet működéséből következő összefüggések szabják meg, melyeket a továbbiakban funkcionális-logikai kapcsolatoknak nevezünk.

A tárolt adatok a gazdasági szervezetek erőforrásai: közvetlenül felhasználják őket az értékelési és döntési tevékenységben. Egy részük változatlanul maradó vagy csak ritkán változó törzsadat, másik – többnyire jelentősebb – részüket viszont a nagyfokú változékonyság jellemzi: ezek az esemény- vagy tranzakciós adatok. Mindezeknek a programrendszereknek nincs még mindenki által egyformán elfogadott, egyezményes nevük. Mi a továbbiakban irányítási célú vagy ügyviteli (vállalati) programoknak nevezzük őket.

Vannak programok, melyek feladata elsősorban ugyancsak az, hogy adatokat tároljanak, és ugyancsak azért, hogy valaminek a megoldását támogassák, de ez a megoldás egy sokkal tágabb és intézményesen korántsem annyira szervezett rendszernek is része, melyet szakmának, tudománynak, ismeretterületnek stb. nevezünk. Amit ezek a programok megoldanak, az lényegében nem több, mint hogy a tárolóból az adott szakmában felhalmozott adatokat nyelvileg érthető formában és összefüggésekben kiadják.

Ezekre az adatokra a nagyfokú állandóság jellemző: ők azoknak a szakismereteknek az alapjai, melyekkel az előző – szervezettebb – rendszerekben az erőforrásként kapott, esetről esetre változó adatokat értékelni lehet. A szakmai, tudományos és egyéb – a kultúra által szervezett – rendszerek beépülnek a civilizáció és a gazdaság rendszereibe, de ugyanakkor át is nyúlnak rajtuk, és keresztül-kasul fonják az egész társadalmat. Ha a szakmai, tudományos és egyéb, hozzájuk hasonlóan csak lassan változó ismereteket rögzítik, dokumentumokról beszélünk. A tárolt adatok vagy e dokumentumokra vonatkoznak, vagy e dokumentumok-



AZ EMBERI - SZELLEMI - BEAVATKOZÁS ÉS MUNKA NÖVEKEDÉSÉNEK MÉRTÉKE

1. ábra A gazdasági, irányítási és kulturális szervezetekben használt főbb programfajták

ból származnak. Az utóbbi esetben a teljes dokumentumok kerülnek gépi hordozóra, vagy rövidített, kivonatos, átdolgozott formájuk. Ezért ezeket a programokat nevezik dokumentációs – vagy könyvtári körülmények között bibliográfiai – célú programoknak. Mivel újabban már az is tapasztalható, hogy a szakismereteket közvetlenül – a hagyományos dokumentumok megkerülésével – rögzítik a géppel olvasható adathordozókon, dokumentációs helyett szaktájékoztatói célú programokról is beszélnek. Valójában ilyenkor nem papíros, film, festővászon stb., hanem mágnesszalag alapú dokumentumokról van szó, hasonlóan a hang- és zenei dokumentumokhoz. Mi a továbbiakban a dokumentációs/szaktájékoztatói és

bibliográfiai programok összefoglaló megnevezésére egységesen a dokumentációs (célú) elnevezést használjuk (indoklásával részletesebben a 2. fejezetben foglalkozunk).¹ A mondottakat az 1. ábrán foglaltuk össze.

¹ A KSH adatfeldolgozási termékek jegyzékében ezeket "szervezetek szakmai tevékenységét ellátó programoknak" nevezik, "melyek a népgazdaság egyes ágainak (ipar, kereskedelem, egészségügy, tudomány, könyvtárügy, igazgatás, közlekedés stb.) tevékenységét segítik elő", szemben a "szervezetek irányítási rendszereinek programjaival", melyek "valamely szervezet gazdasági, pénzügyi, munkaügyi, személyzeti irányításának és gazdálkodásának egyes műveleteit végzik" (KSH, 1982).

1.2 A tárolt adatállományok szerkezete

1.2.1 Az adatbázis és kezelőrendszere

Az adatokat mind a vállalati-ügyviteli, mind pedig a dokumentációs programok esetében olyan rendezettségben tárolják, hogy ez nemcsak a gazdaságos tárolást, hanem az adott igényeknek megfelelő felhasználást (megjelenési, változtatási, keresési stb. lehetőségeket) is biztosítsa. Az adatállomány szerkezetére tehát a mindenkor felhasználati kör "világa" is rányomja a bélyegét: az állománynak nemcsak fizikai, hanem a felhasználó szempontjait tükröző logikai szerkezete is van. E logikailag strukturált, a vállalati erőforrásként vagy szakmai tájékozódásra használható adatállományokat a továbbiakban adatbázisnak, azt a programrendszert pedig, amely létrehozásukat, fenntartásukat, kezelésüket és szolgáltatásaikat biztosítja, adatbázis-kezelő rendszernek nevezzük.

1.2.2 Az adatbázis funkcionális-logikai modellje: a séma

Az irányítási adatbázisok logikai szerkezete és használati módja manapság jelentősen eltér a dokumentációs adatbázisokétól (az eltérés átmeneti jellegrével az 5. fejezetben foglalkozunk). Az első esetben elég pontosan megtervezhetők és formalizálhatók a várható kérdések (a felhasználói kör "világa" jól modellezhető), mert az irányítási-ügyviteli műveletek, funkciók nagymértékben előre láthatók és formalizálhatók.

Egy kereskedelmi vállalat értékesítési és beszerzési rendszerében például előre adott, mi jellemzi a lehetséges vevőt (neve, címe, egyszámú száma stb.) és a megrendelést (dátuma, szállítási címe stb.), továbbá az is, hogy a kettőt mi kapcsolja össze (pl. a szállítási cím — lásd a 4.5 fejezetben a 4. és 5. ábrát).

Úgy is mondhatjuk, hogy a lehetséges kérdések megfogalmazásának formái, szerkezetei előre adóttak. Ezt az előre megadott szerkezetet az adatbázis funkcionális-logikai sémájának nevezzük.

1.2.3 Az adatbázis szemantikai-logikai modellje: a szótár

A dokumentációs adatbázisok esetében rendkívül különböző kérdésekre kell számítani, hiszen mind a szakmai problémák (a témák, tárgyak), mind a rájuk vonatkozó kérdések sok szempont alapján fogalmazhatók meg.

Például a "könyvtári információk automatikus kezelése", a "számítógépes bibliográfiai tájékoz-

tatás", a "dokumentumadatok gépi feldolgozása" és a "szakmai tevékenység támogatásának gépesítése" megfogalmazások lényegében ugyanarra a témára, tárgykörre vonatkoznak.

A felhasználói kör "világát" csak úgy lehet modellezni, ha a lehetséges szavakat és a közöttük lehetséges összefüggéseket figyelembe vesszük. E "világot" tehát maga a szaknyelv modellezi, mellyel a témákat és kérdéseket (problémákat) megfogalmazzák. Ezért az adatbázis logikai szerkezetét nem, vagy nemcsak a funkcionális értelemben vett formalizáltság, hanem a nyelvi megformáltság: a "tartalmi", szemantikai-logikai szervezethez is jellemzi. A tárolt információkat és a lehetséges kérdéseket a szaknyelv szavai és a közöttük fennálló értelmi összefüggések strukturálják.

Az ilyen adatbázisokban kitüntetett jelentősége van a szótárnak, mely nemcsak osztályozási rendszer, tárgyszójegyzék, teaurusz stb. lehet, hanem a címszavak, nevek, lelőhelyek, betegségmegnevezések, tárgy- (objektum-) megnevezések, helynevek stb. állománya is. A dokumentumok vagy a kérdések tartalmának szemantikai-logikai struktúráját elvileg a szótárak hivatottak formalizáltan reprezentálni. Eredményesen — kereshetően — tárolni és kérdezni az általuk adott keretek között lehet.

Ezekben a szótárakban azonban ma még csak a nyelvi rendszer lehetséges modelljének alig töredékét sikerül csak formalizáltan megragadni: a szavakat és rendkívül leegyszerűsített formában néhány fontosabb értelmi összefüggést (szinonimákat, homonimákat, alá- és fölérendelési viszonyokat, aszociációkat), még ritkábban bizonyos mondattani (szintaktikai) összefüggéseket (relátorok, szereplők stb. formájában).

A megragadott szavak nagyon különbözők lehetnek — például személynevek, szakkifejezések, helynevek, cégnevek, részlegnevek, címek, időadatok, nyilvántartási számok stb. —, és a keresés szempontjából elvileg szükségtelen, hogy azonosított helyük és nevük legyen a tárolt szövegben. A teauruszokban, osztályozási rendszerekben, egységesített besorolási állományokban (az ún. "authority fájlokban") a szavak bizonyos fajtáit — például a szakkifejezéseket vagy a szerzők, közreadók neveit — szabványosított formában rögzítik. Az esetek többségében azonban csak a felhasználó fejében, szubjektíven és elég ellentmondásosan van adva e struktúra (lásd a 3. fejezet [d] pontját). A felhasználó az adatbázisban a szavak jelentése — természetes nyelvismerete — alapján tájékozódik. E látszólag kaotikus szabadságra a rugalmas szakmai tájékozódás érdekében ugyanúgy szükség van, mint az egyértelmű döntések érdekében arra a látszólagos "kényszerzubbonyra", mely a funkcionális-logikai szerke-

zet következtében az irányítási rendszerek sajátossága. Két teljesen ellentétes szükséglet kielégítéséről van szó, s ezért nem sikerül ma még a kettőt egyetlen programrendszerben közös nevezőre hozni (a hardver fejlettsége ebből a szempontból másodlagos).

Az adatbázis kezelőrendszere — pontosabban szemantikai-logikai komponense — biztosítja, hogy az adatbázishoz a természetes, beszélt nyelv szavai alapján hozzáférhessenek. Ha szótár van, akkor e szemantikai-logikai hozzáférés pontosabban tervezhető. Ha a szótár strukturált, értelmi összefüggéseket is tartalmaz, a pontosság tovább fokozható. A szótár az az eszköz, amellyel a dokumentum tartalmát, illetve a tartalmára vonatkozó kérdéseket "letranszformálják" arra a szintre, amelyen az adatbázis-kezelő rendszer fizikailag működik. (Azt is meg kell azonban mondani, hogy a legtökéletesebb megoldás, ha e "transzformációt" a szakmát és a nyelvet ismerő szakemberre bízják. Az ilyen szakember már ma is kiképezhető, a vele egyenrangú szabványosított szótár talán soha.)

1.3 Terminológia: bibliográfiai és faktográfiai adatbázis/"igazi" adatbázis-kezelő rendszer és információkereső rendszer

Ha a bibliográfiai és dokumentációs célú adatbázisokat a pályához nem kötött közlekedéshez hasonlítjuk, akkor az irányítási célú adatbázisok a vasúti, kötött pályás közlekedési hálózatot képviselhetik. Az eltéréseket — ma még szakterületenként eléggé eltérő és ellentmondásos módon — a nyelvhasználat is jelzi:

- ◆ Könyvtári és dokumentációs környezetben az első esetben bibliográfiai adatbázisról vagy egyszerűen csak gépi, automatizált nyilvántartásról (dokumentációról), az utóbbi esetben pedig faktográfiai adatbázisról beszélnek — elég szerencsétlenül, mert a bibliográfiai adatbázisban legalább annyi tényadat található, mint a "faktográfiaiában".
- ◆ Irányítási-vállalati környezetben egyszerűen csak adatbázisról beszélnek, mert vagy észre sem veszik, hogy létezik másféle adatbázis (annyira természetes a dokumentáció vagy a könyvtár létezése), vagy fel sem merül, hogy a dokumentációs-bibliográfiai célú adatbázisokat mások is jellemezhetnék, mint a mindennél fontosabb vezetési-irányítási-ügyviteli-vállalati szempontok. ("Szakemberek vagyunk — mit érdekelne bennünket a szakma maga?" — mondhatnák nemes egyszerűséggel, emlékeztük szép-irodalmi "adatbázisából" József Attila *Ars poetica*-ját segítségül hívva.)

Ez persze — tisztelet a kivételnek — a bibliográfiai és dokumentációs szakemberekre sem kevésbé vonatkozik; sőt: nekik éppen feladatuk volna a differenciált megkülönböztetés! Az ő környezetükben mintha az irányítási-ügyviteli kezelőrendszerek nem léteznének — miközben folyton folyóirat-nyilvántartásról, szerzeményezési rendszerről, katalogizáló programról beszélnek, melyek (mint ezt a 6. fejezetben látni fogjuk) inkább irányítási-ügyviteli (vállalati = könyvtári), vagy akár faktográfiai rendszerek, semmint bibliográfiaiak vagy információkeresők.

- ◆ Számítástechnikai környezetben általában csak az általunk irányítási, ügyviteli (vállalati) célúnak nevezett adatállományokat nevezik adatbázisoknak, mert logikai struktúráján csak az adatok közötti funkcionális-logikai struktúrákat értik, nem pedig minden logikai struktúrát (mint amilyen például a helyesen megfogalmazott kijelentő mondat tartalma, vagy egyáltalán: minden természetes nyelven megfogalmazott szöveg). Éppen ezért a dokumentációs és bibliográfiai adatbázisokat szöveges adatállományoknak vagy szöveges adatrendszereknek, vagy -szerkezeteknek nevezik, ami két szempontból is megtévesztő: egyrészt, mert a dokumentációs célú adatbázisok többsége manapság alig tartalmaz szöveges adatokat (a szerző neve, a múzeumi tárgy megnevezése, a tárolás helye, a készítés dátuma, a tárgyszó, az osztályozási jelzet stb. semmivel sem szövegszerűbb adat, mint a beosztás, a munkahely, a végzettség, a születési hely, a szervezeti egység vagy a munkaköri leírás adata). Másrészt, mert sem a szövegek belső, "tartalmi" logikájának teljesen automatikus kezelése, sem a nagy mennyiségű szövegben végzett gazdaságos keresés nem tekinthető megoldottnak (még nagyszámítógépeken sem, mikroszámítógépekről nem is beszélve). Ezért az igazán szöveges adatbázis legfeljebb a jövő zenéje.

A fentiek miatt adatbázis-kezelő rendszernek is csak az irányítási célú rendszerek kezelőprogramjait nevezték. Mivel a dokumentációs és bibliográfiai célú rendszerekben a szabadon vagy szótárból választható szavakkal végzett keresésnek fontos szerepe van, adatbázisaik kezelőrendszerét információkereső rendszereknek szokták nevezni.²

² A kétféle kezelőrendszer kérdését magyar nyelven Halassy (1978, p. 82–83. és 1986) érinti. A külföldi közlemények közül utalunk Tagg (1982) ugyanebben a számban tömörített munkájára.

Az "információkereső" jelző ugyancsak megtévesztő. Szűkebb értelemben ui. csak olyan kezelőrendszert jelent, melynek kizárólagos rendeltetése a keresés és a keresett találatok lehető leggyorsabb és legegyszerűbb megjelenítése. Számátalan dokumentációs célú adatbázis-kezelő rendszer készül ezzel a "filozófiával", és csak a felhasználás során derül ki, hogy mennyire szegényes egy ilyen rendszer (lásd az 5. fejezetet). Tágabb értelemben ui. "információkereső" minden olyan kezelőrendszer, melynek tárgya dokumentum vagy rá vonatkozó információ. Az információkereső rendszer kifejezést használó számítástechnikai szakember szemében tehát egész általánosan egy könyvtári katalógus kezelőrendszere is "információkereső", holott itt a megjelenítési, a rendezési és a különféle egyéb szolgáltatások következtében már nem is lehet önmagában vett keresésről beszélni.

Valójában ilyenkor is adatbázis-kezelésről van szó, csak éppen másfajta adatbázisról, mint az irányítási-vállalati adatbázisok esetében. Tiszteletben tartva mások (pl. *Halassy*, 1978. p. 82-83.) véleményét, mi adatbázison adatok – valamiféleképpen szervezett – bázisát értjük; olyan tárolótartalmat, melyhez felhasználói szándékkal lehet fordulni. Ilyen értelemben akár egy egyszerű fájl is adatbázis lehet – ha adatok szerzése céljából felhasználók igénybe veszik.

A tárolótartalom egyrészt meghatározható úgy, hogy megadjuk azokat az adatfajtákat (adattípusokat, célinformáció-fajtákat – lásd a 4.4 fejezetet), melyek elvileg előfordulhatnak benne, másrészt úgy, hogy megadjuk mindazokat a konkrét értékeket, melyek előfordulnak bennük. E kettősség lényegében ugyanaz, mint ami a fogalom tartalma és terjedelme esetében tapasztalható (*Ungváry*, 1986); a fogalmat megadhatjuk az ismertetőjegyei (= "adatfajták"), illetve a terjedelméhez tartozó dolgok (= "értékek") alapján. Amikor adatbázisról beszélünk, a tartalma alapján meghatározott tárolóra gondolunk (a tároló elvileg lehetséges tartalmaira), amikor pedig adatállományról (fajlról), akkor a terjedelme, az értékei alapján meghatározott tárolóra. Még a legegyszerűbb mai információkereső rendszerekben is meghatározható az adatbázist jellemző tartalom: ti. "dokumentum" célinformációkat tartalmaz.³

³ E ténnyel vet számot *Beling* és *Wersig* (*Beling*, 1973. p. 74-75.), amikor így definiálnak: "Egy rendszer adatbázisa mindazoknak az adatfajtáknak az összessége, melyek a rendszeren belül rendelkezésre állnak." Továbbá: "Egy rendszer adatállománya mindazoknak a tárolt adatoknak az összessége, melyek egy rendszeren belül (konkrétan) hozzáférhetők." Az – elsősorban irányítási-ügyviteli – adatbázisok tervezésére a legnagyobb befolyást gyakorló *Conference on Data System Languages Data Base Task Group* (adatrendszer nyelv-

A gyakorlat mindenesetre az általunk használt és a köznyelvihez közelebb álló értelmezés irányába mutat: a mikroszámítógépek elterjedésével megjelentek a kereskedelmi forgalomban olyan általános programcsomagok (pl. a dBASE vagy az MDDBS, lásd *Schumacher* [1985] ugyanebben a számban tömörített tanulmányát), melyekből a felhasználó megfelelő programozási ismeretek alapján akár irányítási, akár dokumentációs célú rendszert, szükségképpen akár információkereső rendszert is kialakíthat. Az 5. fejezetben ugyan rámutatunk arra, hogy ez a paradicsomi állapot még korántsem realitás, de ez nem változtat azon a tényen, hogy ezeket az általános célú termékeket – tekintet nélkül az alkalmazásukra – ma már adatbázis-kezelő rendszereknek nevezik.

Az adatbázist és kezelőrendszerét együttesen adatbázisrendszernek (*Halassy*, 1978. p. 79. és 105.) nevezik. Amikor a szűkebb értelemben számítógépes (automatizált) információs rendszerről beszélünk, lényegében adatbázisrendszert értünk rajta – vagyis az adatbázist és környezetét, amelybe elsősorban a kezelőrendszere tartozik bele. Ilyen értelemben a dokumentációs adatbázis és kezelőrendszerének (az utóbbit nevezik információkereső rendszernek, szöveges adatkezelő rendszernek, bibliográfiai adatbázis-kezelő rendszernek stb.) az együttese is információs rendszer. (Az információkereső rendszernek persze még további, nem automatizált összetevői is vannak, melyekkel itt nem foglalkozunk.)

Az elmondottak alapján a dokumentációs (célú) adatbázis-kezelő rendszer helyét a 2. ábrán ábrázoltuk összefoglalóan. Ez az ábra ún. címkézett, irányított gráf formájában ábrázolt modellje annak a gondolatmenetnek, melyet e fejezetben leírtunk, azaz rendkívül leegyszerűsítve e szöveg szemantikai, "tartalmi" logikai struktúrája (hasonlóan a 10. ábrához, mely a 6. fejezet szövegének szemantikai struktúráját formalizálja rendkívül leegyszerűsítve). A dokumentációs célú adatbázisban tárolt szöveges vagy egyéb, a dokumentumok tartalmára vonatkozó információk (tárgyszavak, jelzetek stb.) szemantikai szerkezetének lényegében ilyen a "formája".

A "tartalmi" vagy szemantikai struktúra a kifejezések között fennálló különféle kapcsolatokból épül fel, melyeket a tezauszokban szokásos módon kü-

vek bizottságának adatbázis-munkacsoportja definíciójában e kettősségre sajnos nem reflektálnak: "Az adatbázis a felhasználónak megfelelő formában kialakítható, megváltoztatható és kiadható, meghatározott rendszer által fenntartott állományok összessége" (*Codasyl*, 1971). De még e meghatározás sem zárja ki, hogy az információkereső rendszerek által kezelt állományok a felhasználó szempontjából ne legyenek adatbázisok. Eme adatállományok a szemantikai, tartalmi igényeknek megfelelően vannak kialakítva.

lönféle nyilakkal jelöltünk. A folytonos, egyirányú nyíl azt jelöli, hogy az egyik fogalom fajtája a másiknak, a szaggatott azt, hogy része, a pontvonalas azt, hogy eszköze, a kétirányú nyíl pedig, hogy egyéb, nehezen meghatározható okból van összefüggés a két kifejezés között.⁴

Az ábra a következőképpen "olvasandó": az adatbázis-kezelő rendszer programrendszerfajta, része az információs rendszernek, tárgya az adatbázis. Fajtái:

- ◆ a cél, rendeltetés szempontjából az irányítási-ügyviteli és a dokumentációs;
- ◆ a működés, struktúra szempontjából a funkcionális-logikai (a hagyományos, szűkebb értelemben vett adatbázis-kezelő rendszer) és a szemantikai-logikai (az információkereső rendszer);
- ◆ a kezelt információfajták szempontjából a faktográfiai és a szöveges;
- ◆ a felhasználási hely (szervezet) szempontjából a könyvtári, múzeumi, vállalati, személyi stb. adatbázis-kezelő rendszer.

A funkcionális-logikai kezelőrendszer – manapság – szorosan összefügg az irányítási típusú adatbázis-kezeléssel (ezt jelzi a kétirányú nyíl) és a faktográfiai típusú adatbázis-kezeléssel, a szemantikai-logikai pedig a szöveggel. Felfogásunkban a bibliográfiai adatbázis-kezelő rendszer egyik fajtája a dokumentációs: ugyanis dokumentumokra vonatkozó információkat kezel a bibliográfiai szabványok által előírt formában. A könyvtári rendszerben viszont nemcsak bibliográfiai adatbázis-kezelés szükséges, hanem például a beszerzés, a kölcsönzés stb. nyilvántartása is: ezért a könyvtári rendszerek tartalmazzak mind funkcionális-logikai, mind szemantikai-logikai kezelőrészt, s ezért mind irányítási, mind pedig dokumentációs rendszernek felfoghatók (a dokumentációs rendszerek fajtáit a 6. fejezetben foglaltuk össze). A szervezet tehát, melyben egy bibliográfiai vagy egy dokumentációs programrendszert használnak, lehet rendkívül eltérő, de a programrendszer lényegében nem fog különbözni egymástól.

2. Dokumentáció és tájékoztatás

Mielőtt a dokumentációs adatbázis-kezelő rendszerek fajtáival foglalkoznánk, szükségesnek tartjuk tisztázni, hogyan értelmezzük a dokumentáció fogalmát, és miért ezzel a szóval jellemezzük az általunk tárgyalt adatbázis-kezelő rendszereket.

Dokumentáción azt a tevékenységet értjük, melynek során "információkat gyűjtenek és tárolnak, rendeznek és válogatnak, terjesztenek és hasznosítanak" (FID, 1960).⁵ Az információk lehetnek dokumentálatlanok és dokumentáltak, gyűjtésük során rögzíthetik őket hagyományos hordozókon (cédulán, lapokon, füzetben stb.), filmen, hangszalagon vagy géppel olvasható hordozón (mágneslemezen stb.). Terjesztésükre és hasznosításukra ugyancsak az előbbiekre vonatkoznak. A hordozó és a rajta rögzített információ együttesét dokumentumnak nevezzük.⁶

A dokumentáción értjük azonban:

- ◆ nemcsak a fenti tevékenységet, hanem
- ◆ a tevékenység eredményét (a dokumentumok rendezett összességét),
- ◆ a rendszert, melynek keretében a tevékenység folyik, és az eredmény megjelenik,
- ◆ a szakterületet, amelyen e tevékenység folyik.

Ellentétben a teljesen határozott jelentésre törekvő felfogással, ezt a fenti poliszemiát hasznosnak tartjuk – ettől gazdag tartalmilag éppen az a fogalom, amely (hasonlóan az információ fogalmához) az általunk tárgyalt kérdés egyik legáltalánosabb fogalma. Éppen ezért előny, hogy a fogalmat jelölő kifejezés ilyen sokszínű: szükségtelenek a folyamatosan változó, pontos, de nehezen kezelhető, összetett leíró kifejezések – mindig a szövegösszefüggés szabja meg, milyen értelemben használatos a szó.

Az utóbbi évtizedben a hangsúly a gépesítés következtében áttevődött a dokumentumokról az információkra, azaz a dokumentumoktól független szakmai tájékoztatás támogatására.⁷ Ezért a dokumentáció mellett megjelent a (szak)tájékoztatás fogalma. A német szakirodalomban például "tájékoztatásról és dokumentációról" ("Information und Dokumentation") beszélnek (*Grundlagen...*, 1980. p. 6.).⁸

⁵ Valamivel jobban hangsúlyozza a felhasználót az UNESCO meghatározása: "Dokumentumok vagy adatok folytonos és rendszeres feldolgozása, beleértve a felkutatást, azonosítást, beszerzést, elemzést-feltárást, tárolást, keresést, forgalmazást és tartósítást a felhasználók speciális tájékoztatása érdekében" (*Wersig*, 1976. p. 98.).

⁶ Az UNESCO által kiadott értelmező szótár szerint: "Hordozóból, a rajta rögzített adatból és az adat által jelölt jelentésből álló egység" (*Wersig*, 1976. p. 39.).

⁷ Meghatározásában éppen ezt emeli ki *Wersig*: "Az információ és dokumentáció olyan gyakorlati tevékenység (szakterület), amelynek segítségével megkísérlik... a tájékoztatási helyzet javítását" (*Grundlagen...*, 1980. p. 16.).

⁸ A kérdéstről a 60-as évek végén *Koblitz* javaslatai nyomán – aki egyik szószólója volt a kettős kifejezés bevezetésének – vita bontakozott ki, melynek *Polzovic Iván* is résztvevője volt (ennek köszönhetően maradt nyoma az egyébként terminológiai és "filozofikusabb" szempontból végtelenül sze-

⁴ Azt is mondhatjuk, hogy gondolatmenetünk modellezését a teauruszelvek segítségével végeztük el (*Ungváry*, 1983).

Valójában mindkét esetben olyan ismeretek tárolásáról és kezeléséről van szó, melyek tudományos, szakmai vagy tapasztalatilag adott tényeken alapulnak, azaz különböző időkben, különböző helyeken többször is felhasználhatók helyzethez kötött, aktuális problémák megoldásában. Ezeket az ismereteket bizonyos tartósság, állandóság jellemzi, legalábbis nem máról holnapra változnak meg. Eme tartósságuk és ebből következő ismételt felhasználhatóságuk miatt kerül sor dokumentálásukra. A felhasználó nem azért keresi őket, hogy velük kapcsolatban valamilyen intézkedést hajtson végre, hanem azért, hogy tudását gyarapítsa, és adódó alkalommal valaminek a megoldásában alkalmazza őket. Ilyen szakismeretek például az állattenyésztési tudnivalók (melyekről szakkönyvek, szakkikkek számolnak be), vagy az olyan információ, hogy az elfekvő készletek bizonyos nagysága fölött büntető kamatokkal kell számolni (melyet bizonyos gazdasági-pénzügyi dokumentumok tartalmaznak). A dokumentumot teljesen általánosan értelmezhetjük: lehet hagyományos, papíron rögzített szakirodalmi dokumentum, lehet tárgy (pl. régészeti lelet, festmény), akár esemény is, ha valahol dokumentálják, azaz a rá vonatkozó információkat rögzítik. Adott esetben magneses adathordozón is.

Az irányítási, ügyviteli, vállalati információk esetén ezzel szemben olyan ismeretek tárolásáról és kezeléséről van szó, melyek egy adott szervezethez, körülményhez, helyzethez, esethez kötött tényeken alapulnak, azaz meghatározott helyen vagy meghatározott időpontban kell ismerni őket. Ezeket az "ismereteket" a nagyfokú aktualitás, helyhez kötöttség és részben a változékonyság jellemzi. Ilyen információk például a tejhozam csökkenésének adatai egy mezőgazdasági nagyüzemben, vagy az olyan raktári adat, amely az elfekvő készletek nagyságáról tájékoztat. Velük kapcsolatban mindig dönteni kell — ezért tekinthetők irányítási, ügyviteli információknak. A helyes döntéshez azokat az állandóbb szakmai ismereteket kell felhasználni, melyeket az előbbi — vajon milyen nevű? — rendszerekben tárolnak kereshetően.

(Persze dönteni dokumentumokra vonatkozó információk esetében is kell. Például egy katalógus esetében ilyen irányítási döntés, hogy a kezelő mit hova soroljon be, vagy helyes-e egy adott bibliográ-

gényes hazai szakirodalomban). Polzovics szerint a dokumentáció átfogóbb, ennek a szaktájékoztatás csak része. Koblitz szerint a két kifejezés egyenrangú, s a német szakirodalomban hosszú ideig ez a felfogás érvényesült. Ma létezik olyan álláspont, amely szerint a szaktájékoztatás az átfogóbb, mondván, hogy létezik szakmai tájékoztatás, mely nem a szakirodalmi dokumentumokhoz kapcsolódik. (A vitára vonatkozóan lásd *A dokumentáció...* 1961.)

fiai leírás stb. A katalógus *kezelőjének* a döntései pontosan olyanok, mint az irányítási rendszerek *használójának* döntései. A katalógus *használójának* döntései azonban teljesen különböznek ezektől: a dokumentumokra vonatkozó információval kapcsolatban neki elvileg nem olyan döntést kell hoznia, amely eme információfajtára van hatással, hanem eme információ, tudás birtokában egy másik rendszerben kell eldönteni valamit. Az irányítási-ügyviteli és a dokumentációs rendszerek ebből következő egymásra utaltságát jeleztük az 1. ábrán.)

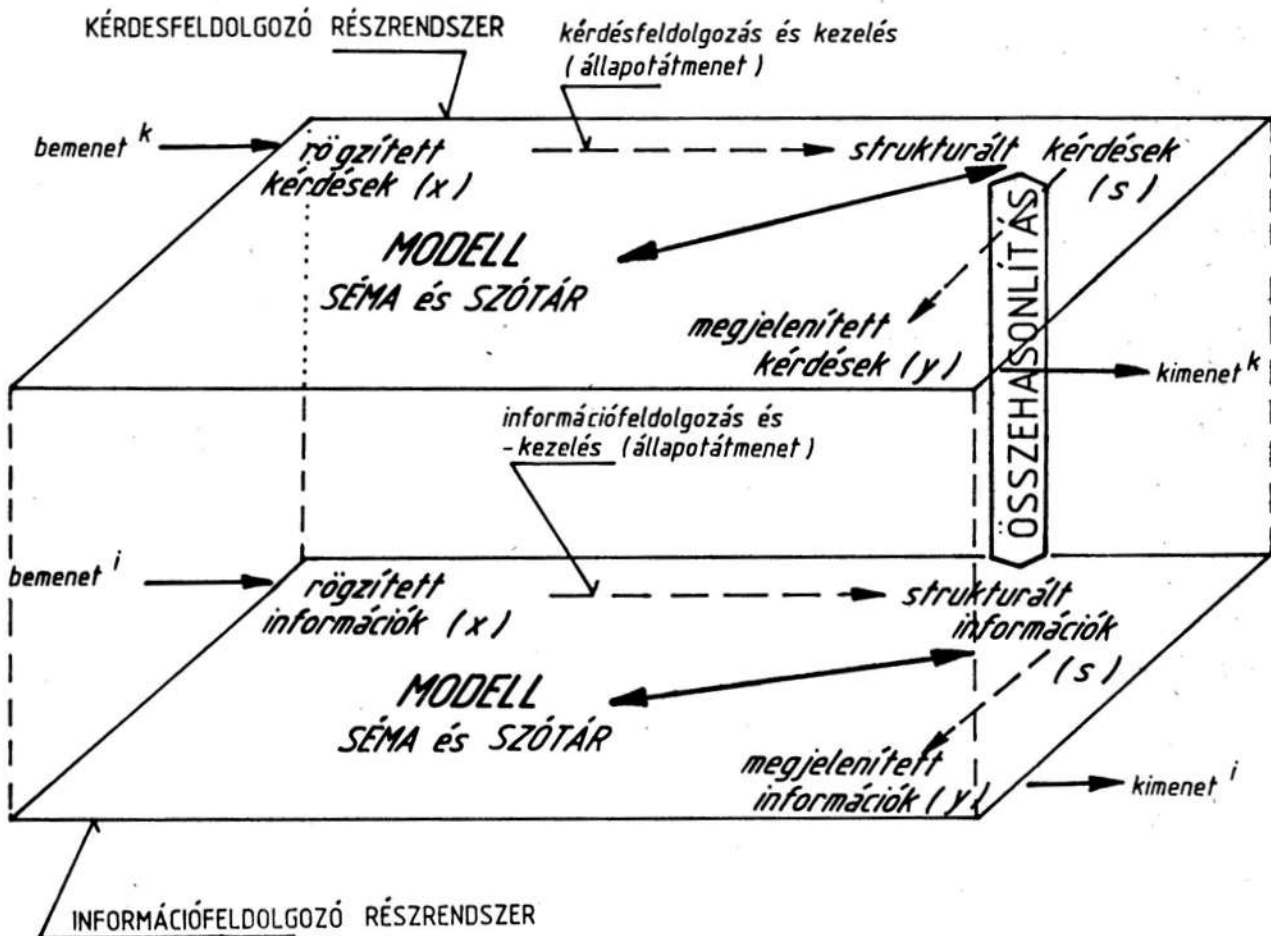
Mindkét rendszer tájékoztatási — azaz információs — célú. E kifejezés egyoldalú használatának tehát a természetes nyelvérzék eleve ellentmondana. Megkülönböztetésükre a következő szembeállítások adódhatnak:



Mi a 2. megoldást választottuk. A tartós, többször is felhasználható szakismeretek feltétele mindenképpen valamiféle dokumentáltság. Nem lényeges, hogy a dokumentumot könyv, festmény, film vagy mágnesszalag képviseli. A dokumentáció és a szaktájékoztatás valójában ugyanúgy nem választhatók el egymástól, ahogy az agy az elmétől, a molekula a vegyülettől, az adat az információtól.

Harmadik megoldásként adódnék az 5. generációs számítógépekkel kapcsolatos "ismeretbázis" ("knowledge base") kifejezés használata (magyarul lásd *Fuchi*, 1983). Így beszélhetnénk "ismeretbázis-kezelőrendszerrel". A dokumentáltságot itt — akár csak az emberi agy esetében — már kizárólag a tár, a memória képviseli. Ez azonban — különösen mikroszámítógépek vonatkozásában — kissé korai volna.

Mindebből természetesen nem következik, hogy ne létezhetnének átmeneti formák is: egy összetett kísérletsorozatot nyilvántartó és kezelő program például dokumentációs és irányítási jellegű egyszerre. Sőt: mint ahogy később még kimutatjuk, maga a dokumentációs adatbázis is tartalmazhat irányítási részt (ilyen pl. a kölcsönzési, szerzeményezési stb.), és elvileg a fordítottja is lehetséges; bár ma még ritka (részletesebben lásd az 5. és 6. fejezetet).



3. ábra A dokumentációs adatbázis-kezelő rendszer általános működési modellje

3. A dokumentációs adatbázis-kezelő rendszer elvi működése

A rendszer működését és felépítését úgy szemléltethetjük, ha absztrakt automatának fogjuk fel (Ungváry, 1984), melynek [a] bemenete (= bemenőalfabéta, X), [e] kimenete (= kimenőalfabéta, Y), [d] kisebb-nagyobb mértékben strukturált adatállománya (= állapota, S) van, és benne meghatározott [d] modell alapján [b] feldolgozás folyik (= állapotátmenet), melynek során egyrészt az adatállomány változik meg a bemenet következtében (SxX leképeződik S -be), másrészt az adatállomány vagy annak bizonyos része a bemenet következtében [f] kiadásra kerül (SxX leképeződik Y -ba) (3. ábra).

[a] A bemeneten a gyűjtés, beszerzés, átvétel, kiválasztás stb. következtében keletkezett adatok rögzítése játszódik le. A bemenőalfabéta (X) a mindenkori ábécé felhasználásával szerkesztett jelsorozatból áll (pl. szerzőnév, cím, referátum, tárgyle-

írás, diagnózis, deskriptor). E jelsorozatok összekapcsolásának és elhelyezésének szabályait mezőnevek, elválasztó jelek stb. formájában a program tartalmazza.

[b] A feldolgozás egyrészt megváltoztatja a beérkező információk tartalmát. Ilyen műveletek az osztályozás, indexelés, annotálás, referálás (kivonatolás), tömörítés, fordítás, az analitikus-szintetikus feldolgozások stb. Ezek automatizálása gyerekipőben jár. Léteznek ugyan automatikus osztályozási, indexelési és kivonatkészítési eljárások, de nem versenyképesek a szellemi feldolgozással. A rendkívül nagy tárigény miatt mikroszámítógépen egyelőre amúgy sem valósíthatók meg. Összefoglalóan *tartalmi feldolgozásnak* fogjuk mindezeket a műveleteket nevezni (nevezhetnénk szemantikai vagy "mélységi" feldolgozásnak is). A szellemi úton végzett tartalmi feldolgozás eredményeként tárgyszavak, deskriptorok, osztályozási jelzetek, címfordítások, tartalmi kivonatok, tárgyleírások, besorolási adatok stb. rögzítésére kerül sor.

[c] A feldolgozás másrészt nem változtatja meg a beérkező információk tartalmát, hanem vagy helyet és nevet ad nekik, vagy megváltoztatja a nevüket és helyüket (strukturálja vagy átstrukturálja az adatállományt). Ilyen műveletek az iktatás, címeírás, katalógusba sorolás, iratrendezés, lelőhely-azonosítás stb. Mindezek a műveletek lényegében azonosítások, rendezések és válogatások; a számítástechnika mai szintjén gyakorlatilag maradéktalanul automatizálhatók. Összefoglalóan *formális feldolgozásnak* fogjuk nevezni őket (nevezhetnénk "szintaktikai" vagy "felszíni" feldolgozásnak is). A formális feldolgozás kiterjed mind a tartalmi feldolgozás során keletkezett adatokra, mind pedig a tartalmilag változatlanul maradó adatokra (pl. iktatási szám, szerzőnév, lacím, megjelenési hely, lelőhely, születési adat, dokumentumcím, foglalkozás, méret). Valójában a formális feldolgozások alkotják az állapotátmeneteket.

[d] A feldolgozás eredményeként kisebb vagy nagyobb mértékben strukturált adatállomány keletkezik, mely a rendszer állapotát (S) képviseli. E struktúra s egyben a feldolgozás alapja — legalábbis az általunk tárgyalt logikai szinten (a fizikai szint kérdéseivel egyáltalán nem foglalkozunk) — annak a környezetnek ("világnak") a modellje, amelyből az adatbázis adatai és a vele szemben támasztott elvárások (kérdések, megjelenítési és átalakítási igények stb.) származnak.

E modell kettős: egyrészt tartalmi logikai — a továbbiakban szemantikai —, másrészt funkcionális-logikai.

A szemantikai modell (lásd az 1.2.3 fejezetet) szótárak (osztályozási rendszerek, tárgyszójegyzékek, teauruszok, címjegyzékek, végzettségkódok, foglalkozásjegyzékek stb.) formájában létezik. A mikroszámítógépek fejlettségének mai szintjén általában nem előre épülnek be gépi szótárként a rendszerbe, hanem különálló szellemi termékek formájában használják őket a tartalmi feldolgozások; a kezelőrendszer csak a ténylegesen felhasznált és bevitt kifejezésekből állítja össze a viszonylag egyszerű betűrendes gépi szótárakat, melyek a további munkában már felhasználhatók. Az is gyakori, hogy egyáltalán nem készül előre szellemi munkával önálló szótár, hanem a feldolgozások eredményeként keletkezett gépi szótárakat nyomtatják ki és használják fel a későbbi munkában szótárként. (A szavak mellett gyakran feltüntetik azoknak a tételeknek a számát is, melyek e szavakat tartalmazzák.) Mint az 1.2.3 fejezetben már utaltunk rá, e szemantikai modell a tárgyszavak, deskriptorok, osztályozási jelzetek stb. esetében valójában sokkal bonyolultabb, mint amennyi manapság még a legigényesebb szellemi munkával készült szótárban vagy osz-

tályozási táblázatban megragadható. A teljes szemantikai modell a (szak)nyelv, a jelentés ismeretének formájában létezik a feldolgozók és felhasználók tudatában. Mintegy "ott lebeg" az adatbázis és kezelőrendszere "fölött", és ebből a bonyolult szemantikai struktúrából csak a szótárakba kerülő szavak épülnek be magába az adatbázisba; ők biztosítják a belépési pontokat (a szemantikai hozzáférést) az adatbázisba. Ritka esetben, ha teaurusz is beépül a rendszerbe, a szavak közötti kapcsolatok — például a 2. ábrán látható formában — valamennyire már strukturálják az állományt. Mindez nem változtat azon, hogy a szótárakba kerülő szavak valóban strukturálják tartalmi szempontból az adatállományt — de ez a struktúra túlnyomórészt nem formalizáltan van jelen.

A szemantikai modellben használt szavak, a szavak közötti kapcsolatok és használatuk szabályainak összességét információkereső nyelvnek nevezik. Az információkereső nyelv szavai a lehetséges szemantikai vagy tartalmi hozzáférési pontokat alkotják az adatbázisban.

Valójában nemcsak a jellegzetesen tartalomra vonatkozó osztályozási jelzetek, tárgyszavak, deskriptorok stb. alkotnak információkereső nyelveket, hanem külön-külön a szerzői nevek, lacímek, helynevek, raktári azonosítók, leltári számok, munkaköri megnevezések, árucikkek nevei stb. is. Ezeket általában nem is szokták önálló szótárakba foglalni, mert hallgatólagosan tudomásul vett szabályok szerint használják őket (például mindenki tudja, hogy a család- és keresztnévet közzel elválasztva, nagybetűvel kezdve, sorrendben előbb a családnévet téve írják). De még ilyen esetekben is előfordulnak szabványosított szótárak, például a folyóiratnevek rövidítései, az országnévek, a címek esetében (ilyen szótárt alkotnak például a mellékletben közölt MSZ 3400, illetve MSZ 3410 szabványokban a nyelv-, illetve országnévek és kódjaik).

A funkcionális-logikai modell az adatbázis sémájának formájában létezik (lásd az 1.2.2 fejezetet). A séma elemeit az ún. adatszótár tartalmazza (*Halassy*, 1978. p. 275.). Szemben az információkereső nyelv szótárával, az adatszótárban nem a konkrét tartalmat esetleg jellemző szavak, hanem e szavak fajtáinak, funkcióinak megnevezései szerepelnek. Az előbbiekre példa: Jókai Mór, Tudományos és Műszaki Tájékoztatás, Vértesszőlős, archeológia, 27 cm (fizikai szinten ezek a mezőtartalmak). Az utóbbiakra példa: Cím, Leleőhely, Tárgyszó, Méret (fizikai szinten ezek a mezőnevek). Ugyanez vonatkozik az összetettebb egységekre is, mint a tételek; így szerepelhetnek az adatszótárban olyan tételfajták, mint "dokumentum" (a dokumentum címét, szerzőjét, megjelenési adatait tartalmazó leírások),

“beszerzés” (az árucikk nevét, a szállítót, az árat, az egyszámlaszámot tartalmazó tétel) stb.

Összefoglalóan azt mondhatjuk, hogy az információkereső nyelv szavai az adatbázis tartalmát *alkotják* (a tartalom elemi értékeit), az adatszótár szavai pedig az adatbázis lehetséges tartalmát előre leírják (a tartalom lehetséges elemi típusait adják meg). Ezért az adatszótár, s vele a séma az adatbázis metaadatbázisa (Halassy, 1978. p. 277.).

[e] A kezelőrendszer edig leírt bemeneti (rögzítési) és feldolgozási műveletei egymással párhuzamosan, két szinten játszódnak le. Az egyik szinten a dokumentált vagy dokumentálandó információk feldolgozása folyik a modellt reprezentáló szótárak és séma alapján: ez az információfeldolgozó részrendszer vagy szint. A másikon az adatbázisnak feltett kérdések és a vele szemben támasztott egyéb igények feldolgozása folyik ugyancsak a modellt reprezentáló szótárak és séma alapján: ez a kérdésfeldolgozó vagy információkereső/lekérdező részrendszer. Mindkét szinten szigorúan ugyanazt a modellt kell alapul venni, azaz ugyanazon a nyelven kell “fogalmazni”, és ugyanazon séma szerint kell a kapcsolatok mentén “haladni” (navigálni). Máskülönben nem lehet a feldolgozás eredményeként strukturált kérdést eredményesen összehasonlítani a feldolgozás eredményeként strukturált adatállománnyal. A kérdéseket például ugyanazoknak az információkereső nyelvi szótáraknak (pl. osztályozási rendszereknek, tárgyszójegyzékeknek, címjegyzékeknek, rövidítésjegyzékeknek) a szavaival kell leírni, mint amelyeket a dokumentumok vagy információk tartalmának feltárásakor használtak. Továbbá ugyanannak az adatszótárnak az alapján kell megadni, hogy az információkereső nyelv szavai az adatbázis szerkezetén belül mire vonatkozzanak (szerzőnevekre, címszavakra, tárgyszavakra, lacímekre, folyóirat-rövidítésekre stb.).

[f] A két szint strukturált állományának összehasonlítása eredményezi a rendszer *kimenetén* a találatokat. A kimenőalfabéta (Y) tehát az információkereső nyelvek és az adatszótár ábécéinek felhasználásával szerkesztett jelsorozatokból áll, az adatbázis sémájának és modelljének megfelelő egységekbe szervezve.

A találatok adott esetben egy tételfajta vagy egy adatfajta teljes állományát is alkotják. A “dokumentum” tételfajta teljes állománya például katalógust alkothat (lásd még a 4.10 fejezetet is), a témát jellemző szavak összessége pl.: a tárgyszójegyzéket.

Elvileg a kérdések is megjelenhetnek a kimeneten, amire van is igény: célszerű ugyanis dokumentálni az alkalmazott keresési eljárást, vagy tárolni a következő alkalomra.

Ha az összehasonlítást a funkcionális-logikai szerkezet, vagyis az adatszótárban megadott szavak és kapcsolatok alapján végzik, lekérdezésről (“inquiry” vagy “query” – Halassy, 1978. p. 432.), ha ebbe az információkereső nyelv szavait és a kereső által megadott logikai kapcsolatokat (halmazműveleteket, Boole-operátorokat stb.) is bevonják, keresésről (“search”) beszélünk. A lekérdezés/keresés legprimitívebb formája a soros, sávban végzett “letapogatás” (“scanning”), azaz a séma ilyenkor csak annyi, hogy “egymás után”. A keresés, lekérdezés és a velük összefüggő műveletek és eljárások összessége az “information retrieval”; nincs rá önálló magyar szó, általában az összetett “információkeresés” kifejezéssel utalunk rá (nem pedig az “információ-visszakereséssel”, mint ahogy nem beszélünk “információ-visszakereső nyelvről” sem).

[g] Mindezek alapján az irányítási és a dokumentációs adatbázis-kezelő rendszer közötti különbség is pontosabban megfogalmazható.

A tiszta információkereső rendszerek nem kezelnek semmiféle funkcionálisan strukturált logikai adatállományt (fizikailag strukturált állomány kezelése persze ekkor is fennállhat). Az adatbázisban nincs a felhasználó szempontjából előre meghatározott mezőnév vagy mezőtartalom, sem különféle tételfajták és közöttük kapcsolatok. (Például a bibliográfiai tételek esetén nem különböztetik meg a tételen belül az egyébként szabványosan kiírt szerzőt, címet, annotációt, hanem funkcionális szempontból egyetlen egységnek tekintik.) A keresés a teljes szövegben, a szemantikai-logikai tartalom alapján történik. Ezek a szövegben szabadon kereső rendszerek (“free text system”): bármelyik szövegben belüli szó alapján a tétel megtalálható. A felhasználói “világ” csak szemantikailag valószínűsíthető, de funkcionálisan előre nem meghatározható. A hozzáférési pontok között csak a szavak jelentése alapján létezik összefüggés. Ezek a rendszerek – akárcsak fejlettebb változataik, az automatikus indexelő és osztályozó rendszerek – ma még elég ritkák.

A tiszta irányítási rendszerek csak funkcionális értelemben strukturált adatállományt kezelnek. Azt is mondhatjuk, hogy a felhasználói “világ” minden igénye és művelete előre meghatározott (szükség-szerű). A belépési pontok között magában az adatbázisban előre megadott, funkcionális kapcsolatok vannak, melyek a megjelenítés eredményébe is szervesen beletartoznak. (Például megadják, hogy az adott beszerzési tétel ára milyen egyszámlaszámra fizetendő be, vagy melyik olvasónál milyen könyv található.) Az ilyen adatbázisok lekérdezésekor csak az előre meghatározott összeállítású kérdések tehetők fel (a szavak mellett a “szintaxis” is kötött). Ezek a rendszerek ma elég gyakoriak.

A gyakorlatban előforduló dokumentációs adatbázis-kezelő rendszerek mindig tartalmaznak olyan összetevőket, melyekkel funkcionális-logikai struktúrák alakíthatók ki az adatbázisban. Mennél kevésbé szakmai, általános érvényű, hanem aktuális, meghatározott döntést vagy felhasználást igénylő információkat kezelnek, vagy megjelenítési formátumokat szolgáltatnak, annál nagyobb mértékben.

Ez a kérdés átvezet a dokumentációs adatbázis-kezelő rendszerek fajtáihoz. Mivel ezek — elsősorban felhasználói szempontból — az általuk kezelt adatok/információk fajtái alapján térnek el egymástól, előbb velük foglalkozunk.

4. A dokumentációs adatbázisokban előforduló adatok/információk a felhasználó szempontjából

4.1 Az adat és az információ kettőssége

Mielőtt az adatok/információk fajtáinak tárgyalására rátérnénk, tisztáznunk kell az adatnak és információnak egymáshoz fűződő kapcsolatát, s azt, hogyan használjuk őket a továbbiakban. E két kifejezés ugyanis egymástól elválaszthatatlan, mégpedig azért, mert az általunk jelölt két fogalomnak ugyanaz a terjedelme, de nem ugyanaz a tartalma; más szóval a két kifejezés ugyanazt a dolgot jelöli (azonos a denotátumuk, a jelöletük), de különböző értelemben teszik ezt. Jól érzékelhető ez a következő példában:

A 19 mint grafikus jel vizuális, érzéki "adat", de ehhez képest az az ismeret, hogy ez szám, információ; a 19 mint szám adat, de ehhez képest az az ismeret, hogy ez hőmérsékleti érték, információ; a "19 fokos hőmérséklet" adat, de ehhez képest az az ismeret, hogy ez a víz hőmérséklete, információ; a "19 fokos a víz hőmérséklete" adat, de ehhez képest az az "ismeret", hogy ez valakinak a fürdéshez még hideg (vagy már meleg), információ.

A skála láthatóan az értelmi szint alatti érzéki szinttől az értelmi szint feletti érzelmi szintig terjed (ezért szerepel az első "adat" és az utolsó "ismeret" idézőjelben), s közte a végtelenségig finomíthatjuk a tisztán logikai kettőződéseket. Hasonló a helyzet, mint az adatbázis és az adatállomány kettőssége esetében (lásd az 1.3 fejezetet és a 3. lábjegyzetet): ha információról beszélünk, a tartalom alapján is meghatározott dologra gondolunk (a dolog lehetséges tulajdonságaira), amikor pedig adatról beszélünk, akkor az előbbi dologra, de csak a terjedelme (a dolog "értékei") alapján meghatározva.⁹ Az információ az ismeret szerepére, hatására, tartalmára utal, az adat pedig az ismeret pusztá tényére, meglétére, formájára. Ezért az előbbi valamivel szorosabb-

an kapcsolódik az ismeretek felhasználásához, az utóbbi pedig az ismeretek kezeléséhez. Talán főképp ennek tulajdonítható, hogy a dokumentáció, a tájékoztatás, a könyvtárügy stb. területén gyakori az előbbi használata, a számítástechnikában — és általában a formálisabb, mechanikusabb megközelítések esetén — viszont túlnyomórészt az utóbbi honosodott meg. Ezzel magyarázható az is, hogy ha általánosabban közelítik meg a kérdést, a számítástechnikában is információról beszélnek (pl. azt mondják: információs rendszer, információkeresés), ha pedig szabványosításról, nagyon is speciális, részletező megközelítésről van szó, a dokumentációban stb. is inkább az adat kifejezést használják (pl. bibliográfiai adatelem, kölcsönzési adatok).

A továbbiakban elsősorban az információ kifejezést fogjuk használni, ezzel is hangsúlyozva, hogy mondanivalónkat egyrészt — noha a számítástechnikai szakembernek is címezzük — a felhasználó szempontjából, másrészt pedig a lehető legáltalánosabban igyekszünk megfogalmazni. A két kifejezésen azonban mindig ugyanazt értjük.¹⁰

⁹ Az ilyen fogalmi kettőzések nem is olyan ritkák. Például a "molekula—vegyület", "atom—elem", "agy—elme", "ismeret—tudás", "civilizáció—kultúra" stb. esetében is páronként ugyanarról van szó, de más-más értelemben (lásd Ungváry, 1986).

¹⁰ Ez a megoldás nincs összhangban számos szabványosítási törekvéssel, melyekben kizárólag az adat használatát szorgalmazzák, mivel az információ fogalmát — Wiener nyomán — az anyag és az energia fogalmaival egyenértékű általánosításnak ismerik el (Beling, 1973, p. 13.). A javasolt definíciókból azonban rendre kiderül, hogy az egyik a tartalmibb, a másik a formálisabb oldalát emeli csak ki ugyanannak a dolognak. Az információ "a határozatlanság csökkentése kommunikációs folyamatok következtében", illetve "olyan adat vagy kijelentés, melyről ... feltehető, hogy tájékoztató hatása van" (Beling, 1973, p. 17. és 22.), továbbá "közölt tény", "tény vagy fogalomhordozó és jel (adathordozó és jelentés) együttese által megvalósított reprezentálásra való üzenet", "adatnak tulajdonított jelentés" (Wersig, 1976, p. 74.). Az adatok "tények legkisebb megvalósítható reprezentációi, melyek ... interpretálhatók és diszkrét jelek formájában zartosan rögzíthetők" (Beling, 1973, p. 21.), továbbá "tények, fogalmak vagy utasítások formalizált reprezentációi..." (Wersig, 1976, p. 72.). Az információ tehát "csökkentés", "hatás", "tény", "üzenet", az adat pedig "jelek formájában megvalósított reprezentáció", "formalizált reprezentáció". Hasonló következtetésre jut Halassy (1982, p. 22.) is: "Az adat ... reprezentált, de nem értelmezett ismeret" és "Az információ ... értelmezett ismeret". Végző soron tehát mindig ugyanannak a dolognak formaibb, illetve tartalmibb megnyilvánulására lyukadunk ki. Ebből az is következik, hogy elsősorban a mindenkori szövegösszefüggés szabja meg, melyik kifejezés használata találhatóbb.

4.2 Eredeti és származtatott információ

A dokumentáltság szempontjából megkülönböztetjük a közvetlenül a dokumentálatlan dolgokra (tényekre, tárgyakra, személyekre, helyzetekre, állapotokra, tulajdonságokra, esetekre, eseményekre) vonatkozó eredeti (elsődleges, primer), a dokumentumokban rögzített információkra vonatkozó másodlagos (szekunder), az ezekre vonatkozó harmadlagos (tercier) stb. információkat. (Értelemszerűen ez a dokumentumokra is vonatkozik: az elsődleges, másodlagos, harmadlagos stb. információkat tartalmazó dokumentumokat elsődleges, másodlagos, harmadlagos stb. dokumentumoknak nevezzük.)

A dokumentációs célból feldolgozott információk lehetnek akár elsődlegesek (pl. a tárgy-, hang-, kép-, személyi, munkaügyi, egészségügyi, gazdasági, határidő-információk), akár másodlagosak (pl. a bibliográfiai információk), természetesen harmadlagosak stb. is. A másodlagos információk azonban nemcsak a hagyományos könyvtári dokumentumokra vonatkozhatnak. Mivel a statisztikai apparátusok manapság már olyan adattömeget gyűjtenek be és tárolnak, melyek a maguk elsődleges formájában áttekinthetetlenek, kialakulófélben van a rájuk vonatkozó metainformációk rendszere, melyek segítségével a statisztikailag rögzített elsődleges információk leírhatók (*Domokos*, 1983). Ezek tehát nem bibliográfiai jellegű másodlagos információk. Nem bibliográfiai jellegűek az irányítási (célú) adatbázis-kezelő rendszerek metaadatbázisaiban szereplő adatszótárak információi sem (*Halassy*, 1978, p. 275.).

A másodlagos információk értelmezésünkben információkat azonosítanak, jellemeznek, ábrázolnak, *irnak le*. A harmadlagosakra ugyanez érvényes a másodlagos információkra vonatkozóan, ezért dokumentációs szempontból lényegében már nem különböznek egymástól. Mind közvetett, származtatott (a hivatkozás, a nyilvántartás, a dokumentációs célú jellemzés és felhasználás vagy a metanyelvi leírás érdekében keletkezett) információk.

Az elsődleges információknak létezik egy másik értelmezése is, amely az információ eredetiségét általában, a dokumentáltságtól függetlenül ragadja meg. Így kritérium pl. az információ újszerűsége (*Lexikon des Bibliothekswesens*, 1974, p. 699.)¹¹ Az újszerűség e feltételének fennállása egyrészt jóval nehezebben ellenőrizhető, mint az, hogy az infor-

máció a dokumentációs feldolgozás tárgya volt-e, vagy sem. Másrészt a számítástechnikai feldolgozás szempontjából nincs jelentősége, hogy valaki mit tart újszerűnek, és mit nem; annak viszont van jelentősége, hogy dokumentációs szempontból feldolgozott vagy feldolgozatlan információkra vonatkozó információk kezeléséről van-e szó, mert hozzájuk teljesen más feldolgozási és megjelenítési igények fűződnek.

Ezért a továbbiakban az elsődleges információk helyett mindig *eredeti információkról*, a másodlagos, harmadlagos stb. információk helyett pedig *származtatott információkról* fogunk beszélni. Az elsődleges információkat a számítástechnikai gyakorlatban alapadatoknak nevezik (*Halassy*, 1978, p. 22.).

4.3 Tény- és szöveges (értelmezett és értelmezésre szoruló) információ

Az információk egyértelműsége, feltártsága, magyarázatot, értelmezést igénylő, illetve nem igénylő mivolta alapján megkülönböztetjük a faktográfiai vagy *tényinformációkat* és a *szöveges információkat*.

A tényinformációk már megjelenési formájukban egyértelműbben kifejeznek egy meghatározott – mondhatnánk azt is, hogy egységnyi – tényállást (pl. "Beosztás: főelőadó", "falvastagság: 5 cm", "Szerző: Gyulai Gergely Ákos", "Főzni jó (Monspart Éva – Bp., Móra, 1981)"); azaz nincs szükség magyarázatra, interpretációra. A tényállások ebben az értelemben azt fejezik ki, hogy egy dolognak milyen összefüggésben, értelemben van valamilyen tulajdonsága, vagy milyen értelemben áll fenn a dolgok között valamilyen összefüggés.

Valami akkor – és olyan mértékben – tényinformáció, ha – és amennyiben – az információ helye vagy egy rá vonatkozó azonosító (név) megmondja, hogy miről van szó.

Ha az adat *helyéből* tudjuk, hogy ő micsoda, akkor pozicionált adat, ha valamilyen azonosító (név) mondja meg, akkor névvel azonosított, megnevezett adat.

Az információnak azt a részét, amelyet így azonosítanak, információtartalomnak (fizikai szempontból mezőtartalomnak), azt a részét pedig, amely azonosít, információnévnek (fizikai szempontból mezőnévnek) nevezik. Valójában megnevezetlen – szöveges – és megnevezett – tény- – információ esetében ugyanaz a kettősség alakult ki az adatfeldolgozásban, mint amelyet az adat/információ kettőssége képvisel a természetes nyelvben (lásd a 4.1 fejezetet).

Mezőnév például az, hogy "Beosztása:", mezőtartalom pedig az, hogy "főelőadó". A mezőnév és

¹¹ "Egy primer dokumentum ... a tudományos munka eredményeként keletkező (eredeti) információforrás (pl. monográfia, szakkikk, disszertáció, szabadalmi leírás), mely új tudományos felismeréseket vagy ismert gondolatok és tények új értelmezését, magyarázatát tartalmazza..."

mezőtartalom között mindig szerepel valamilyen elválasztó jel (interpunkció), példánkban ez a “:”, vagyis a kettőspont és a köz. Ha pozicionált adatokról van szó, az esetek nagy részében a hagyományos értelemben vett, természetes nyelven kifejezett mezőnév el is tűnik, csak az elválasztó jelek maradnak meg. Ilyenkor a mezőtartalom sorrendi helye – és olykor az elválasztó jel fajtája – játssza az értelmező szerepet, amelytől az információ tényinformáció lesz. Ilyen például a fenti példák közül a “Főzni jó...” című könyv leírása. Itt elsősorban a pozíciók alapján tudjuk, hogy címről, szerzőről, megjelenési helyről, kiadóról, megjelenési évről stb. van szó. A pozíció szerepét ugyan ismerni kell – a mező itt nincs természetes nyelven kifejezett névvel jellemezve –, de ez a mesterségesebb megoldás sok esetben általánosabb azonosító erejű: a fenti cím leírása abból a szempontból, hogy milyen értelemben vett adatokról, tényekről van szó, bármely nyelven egyformán érthető volna.

A szöveges információk ezzel szemben olyan – túlnyomórészt sorfolytonos – leírások, melyeket a természetes, beszélt nyelv szabályai szerint tagolnak egységekbe, így mondatokba stb. (pl. “Nagy Ferenc főelőadó”, “Az 5 cm-nél kisebb falvastagságok kétségbeesztő helyzetet idéztek elő”, “Főzni jó”). Mindezeket az információkat a természetes nyelv ismerete alapján megértjük ugyan (előfordulhat azonban, hogy ez sem áll fenn: “A szuperszimmetria megszűnésével kezdetét vette az elemi tömeggel nem rendelkező részecskék leválása”), de az értelmezés ezen túlmenően nem lehetséges. Nem tudhatjuk, milyen összefüggésben áll fenn az adott jelentés, hogy értelmezendő az összefüggés (pl. az “5 cm-nél kisebb falvastagság” és a “kétségbeesztő helyzet” között, vagy a “Főzni jó” nevelő szándékú, ösztönző unszolás, örömteli felkiáltás a siker élménye következtében, vagy esetleg könyv címe). A “Nagy Ferenc főelőadó” esetében látszólag egyértelmű a helyzet. De ha már az állna itt, hogy “Nagy Ferenc hóhér”, nem biztos, hogy a beosztásáról lenne szó, az is lehetséges, hogy valakinek a lelkületét akarják így jellemezni.

Abból, hogy az információ egésze tényinformáció, még nem következik, hogy a részei is azok. Egy annotáció vagy egy referátum – annak alapján például, hogy a cím leírása után következik – egészében tényinformáció (tudjuk ui.: e szöveg értelme, hogy referátum), de a referátum egyes állításai már szöveges információk, melyeket a műveltség, a szakismeret vagy egyéb szellemi-lelki adottságok alapján kell értelmezni. Vagy ha a feldolgozóskor külön kell kezelni a keresztnéveket (mert el kell hagyni őket, vagy másképp kell megjeleníteni őket stb.), akkor “Szerző: Gyulai Gergely Ákos” tény-

adaton belül a keresztnév már nem tényadat, hanem szöveges adat: nem egyértelmű, hogy a Gergely Ákos vagy az Ákos a keresztnév. Ha viszont a keresztnév előtt egyezményesen a vessző és köz elválasztó jelek szerepelnek, a keresztnév is tényadat: “Szerző: Gyulai, Gergely Ákos”. A statisztikai feldolgozás szempontjából például a “Falvastagság: 5 cm” tényinformáción felül külön az “5” és a “cm” már nem tényadatok.

Egy szöveges információ – pl. önmagában egy referátum vagy egy cím szövege, de akár egy szó is – értelmezéséhez már csak az a – dokumentációs szempontból sokkal bonyolultabb – szerkezet (szövegösszefüggés, kontextus) járul hozzá, amely egyrészt a természetes nyelv grammatikáját jellemzi, másrészt az a még bonyolultabb viszony, amely a szöveg értelmezőjét, olvasóját a szöveg tárgyához, jelentéséhez fűzi. A szöveges információk tehát sokkal bonyolultabb szinten tények, éppen ezért nevezük őket szöveges információknak dokumentációs szempontból. Valójában valamilyen szempontból értelmezett (tény-) és nem értelmezett (szöveges) információkról kellene beszélni.¹² A természetes nyelv szempontjából például éppen az adat (értelmezendő ismeret) és információ (értelmezett ismeret) képviselnek hasonló kettősséget.

A tényinformációk egy része közvetlenül a dolgokra vonatkozik (a vállalati, gazdálkodási, egészségügyi stb. adatok), egy másik része viszont a dokumentumokra. Az utóbbiakkal szemben különleges megjelenítési követelmények merülnek fel, melyek alapvetően eltérnek az előbbiekkal szemben támasztott követelményektől, ahol a tablóyszerű, úrlapszerű, táblázatos megjelenítés az uralkodó. E két-fajta követelményrendszer azonban egyetlen szoftver keretében kielégíthető.

A szöveges információk teljes értékű információkereső kezelése – az automatikus osztályozás, illetve indexelés – ezzel szemben teljesen más és rendkívül bonyolult (jórészt megoldatlan) problémákat vet fel. Ezek a mikroszámítógépek mai szintjén még nem kezelhetők, s ezért a továbbiakban az automatikus osztályozás, illetve indexelés kérdéseivel nem foglalkozunk.

4.4 A célinformáció

Az egységként kezelt információk a tételek. A keresés, a felhasználó szempontjából az információtételek vagy a keresés tárgyai, vagy a keresés tárgyának a jellemzői (tulajdonságai).

¹² Önmagában az is szöveges információ a dokumentáció szempontjából, hogy 2x2, ha önmagában csak azt jelenti, amit így jelent, viszont tényinformáció, ha úgy jelenik meg, hogy “alapterület (m²-ben): 2x2”.

A felhasználói keresés tárgyait alkotó információ-tételek a célinformációk.¹³ Ezek mindazok a tételek, melyek a keresés (a felhasználó igényei) szempontjából alkotnak valamilyen egységet. A tároláskor általában nem ilyen egységekbe szervezve kezelik az állományt, ha tehát az egységekről tárolási, számítástechnikai — főleg pedig fizikai — szempontból van szó, a "rekord" kifejezést használják.

A célinformációk azokat a dolgokat (egyedeket, entitásokat), illetve tulajdonságaikat reprezentálják, melyekre a felhasználók figyelme irányul, melyek a "világukra" jellemzőek, s velük kapcsolatban dönteniük kell, vagy a döntésekhez szükséges ismereteket kell megszerezniük. Az első célt általában az irányítási-ügyviteli, a második célt általában a dokumentációs adatbázis-kezelő rendszerek támogatják.

4.5 Irányítási-ügyviteli célinformációk

Az irányítási-ügyviteli adatbázisokat használók általában különbözőfajta dolgokat, szükségképpen különbözőfajta célinformációkat keresnek. Például "vevő", "megrendelési", "cikk", "tétel" célinformációkat (4. ábra), melyek tartalma részben át is fedheti egymást. Az egyes célinformáció-fajták között az ügyvitel menetét kifejező és a döntéseket támogató kapcsolatokat is meghatároznak (például a "vevő" a szállítási címen keresztül a "megrendelési", s azon keresztül a "tétel" célinformációhoz kapcsolódik, amely a "cikk" célinformációval kap-

csolatban áll) (5. ábra). Arról van szó, hogy több különböző felhasználó használhatja egyszerre ugyanazt az adatbázist, akiket különböző célinformáció-fajták érdekelnek. Hogy ez a használat integráltan valósuljon meg, ugyanakkor a tárolás és a kezelés se legyen gazdaságtalan, a különböző érdekeltségű felhasználókat tükröző célinformációfajták között előre megadják az összefüggéseket. (A számítástechnikában sokszor csak ezeket tekintik "igazi" adatbázisoknak (Halassy, 1978, p. 82–89.); ha ez a logikai szerkezet hiányzik, inkább fájlkezelő vagy adatkezelő rendszerekről stb. beszélnek. A célinformáció-fajta helyett a számítástechnikában a rekordtípus vagy az adattételtípus kifejezés használatos.

Ezeket a célinformációkat továbbá az 1. fejezetben már említett tulajdonság jellemzi: elsősorban az adott körülményekhez, helyzetekhez, esetekhez kötött, részben változó adatokat képviselnek vagy tartalmaznak. Az adatok a kereskedelmi, vállalati, gazdálkodási, ügyviteli munka területeiről származnak, viszonylag tömörek, egyértelműek, gyakran számszerűek, tehát olyan tényinformációk, melyeken belül semmi szükség nincs szövegen belüli keresési lehetőségre; alig fordulnak elő bennük szöveges információk. Ezért az adatok mérete jól meghatározható, előre tervezhető, s nem gazdaságtalan az állandó mezőhosszúságok használata.

Jellemző ezekre a célinformációkra, hogy a felhasználói körük — bár lehet nagyon különböző érdekeltségű — nagyon szűk: lényegében egy-egy szervezeti feladat végrehajtásában érdekelt. Az egyes felhasználó mindig ugyanazt az előre meg-

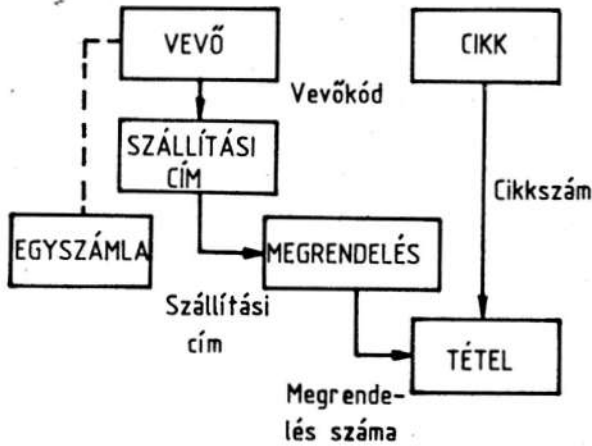
CIKK <Név, Egységár, Mennyiségi egység, Cikkszám>
 MEGRENDELÉS <Szám, Dátum, Vevő azonosítója, Szállítási cím>
 VEVŐ <Neve, Azonosítója, Címe, Egyszámlaszám, Fizetés módja>
 TÉTEL <Cikkszám, Megrendelt mennyiség, Szállítási határidő, Csomagolás módja>

4. ábra Néhány, az irányítási-ügyviteli adatbázisokra jellemző célinformáció-fajta
 Halassy (1978, p. 378.) nyomán

adott, az ügyvitel, a döntés szempontjából gyorsan áttekinthető megjelenítési formát használja, amely többnyire táblázat vagy valamilyen űrlap, s ezek a formátumok a rendszerben előre adóttak. A szűk, specializált felhasználói kör, az előre látható igények, lényegében tehát a nyilvános felhasználás hiánya következtében ezeknek a rendszereknek többnyire rendkívül szegényes a nyelvi komfortja (nehezen megtanulhatók, a rendszerrel folytatott párbeszéd a természetes nyelvtől idegen, azt kerébe töri, és a legkisebb változtatás is a programozó beavatkozását igényli).

¹³ A célinformáció kifejezés a GOLEM szoftvertermék nyomán terjedt el, melyben a 70-es évek elejétől kezdve használták. A KGST-adminisztrációban az "elérési formátum" kifejezést is használják erre a célra: "az adatbázis elérési formátumán a konkrét adatbázis dokumentumának ábrázolása, a keresési üzemmódban lehetséges hozzáférési módok és az információ terminálon való megjelenítése, a dokumentum bontása adatelemekre, az egyes adatelemek jellemzői és különböző üzemmódban lehetséges hozzáférésük (sic!) értendő" (Gornostaev, 1983). E meghatározás kissé szerencsétlen, mert ha a formátum — az értelmező szótár szerint — "megszabott alak", akkor nem lehet egyben hozzáférés, azaz tevékenység.

Egyszámlaszám



5. ábra Egy irányítási-ügyviteli adatbázis jellegzetes logikai adatmodellje Halassy (1978, p. 281.) nyomán. A célinformáció-fajták magyarázata a 4. ábrán szerepel. A funkcionális kapcsolat alapját a nyilak mellé írt azonosító ismérvek képviselik

4.6 Dokumentációs célinformációk

A dokumentációs adatbázisokat használók lényegében egyetlenfajta – vagy csak nagyon kevés különbözőfajta – dolgot, szükségképpen célinformációt keresnek. Ez elsősorban a “dokumentum” célinformáció. Az összes többi fajta – ha egyáltalán van – szorosan ehhez kapcsolódik, csak a “dokumentum” célinformáció léte következtében van jogsultsága. Ilyen célinformáció-fajták például a szelektív információterjesztő rendszerek esetében a “felhasználói”, a kölcsönzési rendszer esetében az “olvasói”, a szerzeményezési rendszer esetében a “szerzeményezési” stb. (6. ábra).

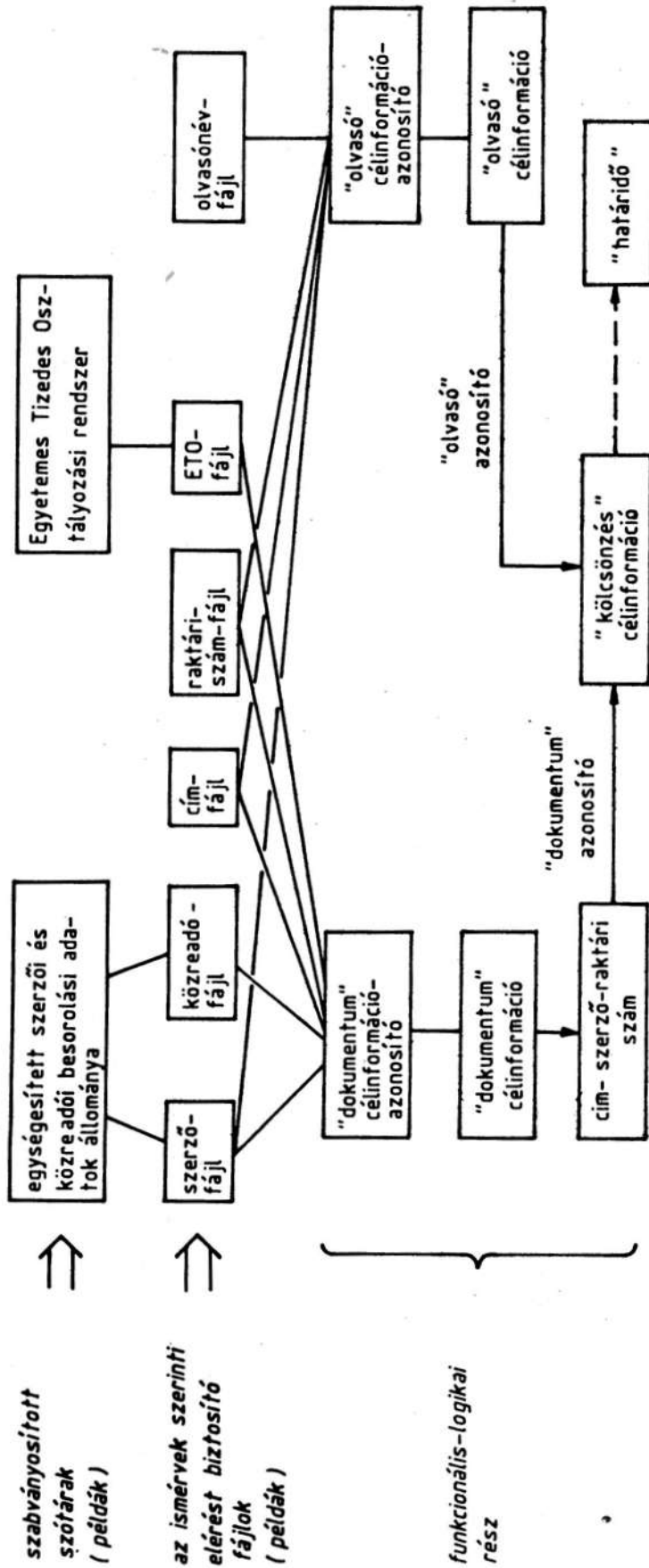
Ezek tartalma főleg a dokumentum vonatkozásában fedi át egymást (többnyire mindegyik tartalmazza a “dokumentum” célinformáció többkevesebb adatát). Az irányítási-ügyviteli adatbázisokra jellemző kapcsolatok e célinformáció-fajták között is megvannak, még ha nem is mindig annyira összetettek a kapcsolataik.

- “DOKUMENTUM” < Cím, Szerző, Megjelenési hely, Kiadó, Megjelenés, Lapszám, Méret, Ár, ISBN-szám, Megjegyzés, Tárgyszavak, Tartalmi kivonat >
- “OLVASÓ” < Név, Lakcím, Munkahely, Végzettség, Kölcsönözhető dokumentumok száma, Megállapítandó határidő, Büntetési kategória >
- “KÖLCSONZÉS” < Cím, Szerző, Raktári szám, Határidő >

6. ábra Néhány, a dokumentációs adatbázisokra jellemző célinformáció-fajta

A “dokumentum” célinformáció jellegzetessége, hogy az egyes tételek tartalma jelentős mértékben átfedi egymást. Az azonos szerzőjű, kiadójú, megjelenési évű könyv, folyóiratcikk, illetve az azonos leltőhelyű, anyagú, kikészítésű régészeti lelet vagy iparművészeti alkotás, az azonos betegséget, gyógyszer, tüneteket tartalmazó lelet e rendszerekben tipikus – és éppen e tény az alapja a halmazműveletekkel végezhető – formális-logikai – kereshetőségüknek. Rendkívül fontos, hogy az ilyen megegyezések alapján, az egyes célinformáció-tételek megtalálhatók legyenek, teljesen függetlenül attól, hogy ki a felhasználójuk, és minek az érdekében igényli őket. Összességük alkotja a 3. fejezet [e] pontjában említett információkereső nyelvet. Különösen kiemelkedő jelentősége van a “dokumentum” célinformációkon belül a dokumentum “belső” tartalmát, témáját, tárgykörét, értelmét, jelentését és jelentőségét feltáró elemeknek, az osztályozási jelze-

teknek, tárgyszavaknak, deszkriptoroknak stb. Rendkívül nagy azoknak a tételeknek a száma, melyek ugyanazt a tárgyszót, jelzetet, deszkriptort tartalmazzák. Elvileg a célinformáció-tételek minden egyes eleme alapján meg kell tudni találni a tételt, és az se mondható meg előre, hogy az egyes elemek mely kombinációira kell számítani a kereséskor. Ezért biztosítani kell a kereséskor egyrészt a 3. fejezet [f] pontjában említett halmazműveletek – a logikai operátorok: ÉS, VAGY, DE NEM – használatát, és azt, hogy a velük végzett keresés gazdaságos legyen. A célinformációk kereshető elemeiből ezért ún. invertált fájlokat alakítanak ki (vagy egyéb módon biztosítják a gyors hozzáférést). Ezeknek nemcsak az olyan elemek, mint szerző, tárgyszó stb., hanem adott esetben akár a címben vagy a tartalmi kivonatban szereplő szövegszó is előfordulhat. Mindez az irányítási-ügyviteli rendszerekéhez képest sokkal nagyobb kárigénnyel jár.



7. ábra Katalógussal összekapcsolt kölcsönzés-nyilvántartási rendszer modellje. Az ábra célja elsősorban az 5. ábrán látható irányítási-ügyviteli adatbázis modelljével való összehasonlítás; nem minden lehetséges elemet (fájl, szólisták) tüntetünk fel

A 7. ábrán gépi – online – katalógus és kölcsönzés-nyilvántartás integrált adatbázisának logikai modellje látható. A "dokumentum" célinformáció-típus bizonyos funkcionális-logikai kapcsolatban van az "olvasó" célinformáció-típussal; ezek strukturált állományai alkotják a funkcionális-logikai adatbázisrészt.

A kereséshez használható szavak és egyéb elemek állományai alkotják az adatbázis tartalmi-logikai (szemantikai) részét, így a név és cím szerinti katalógus egységesített besorolási adatainak¹⁴ ("authority fájl") és az Egyetemes Tizedes Osztályozás jelzeteinek az állományai, valójában azonban a szótárba nem foglalt raktári lelőhely-azonosító számok, az olvasónevek állománya is; és ha úgy döntöttek volna, akár a kiadók és közreadók nevének, az ISBN-számoknak, az olvasók lakcímének stb. állományai is. (Az általánosított adatbázis szerkesztését az 5. fejezetben tárgyaljuk.)

Jellemző a "dokumentum" célinformációkra, hogy viszonylag állandóak, jelentős részük szakismeretekhez kötődik. Akár bibliográfiai adatokról, tárgyszavakról, osztályozási jelzeteiről, akár egészségügyi lelet-, régészeti tárgy- vagy engedélyezett növényvédőszer-nyilvántartásról van szó, a nyilvántartott adatok a dokumentációs adatbázis élettartama során többségükben változatlanok maradnak (sem a dokumentum szerzőjének neve, sem a betegnek az adott kezeléskor felírt gyógyszerei, sem egy adott növényvédő szer tulajdonságai nem változó adatok). Ugyanakkor az adatok egy részének (például a címnek, a referátumnak vagy a régészeti lelet leírásának) a hosszúsága rendkívül változó, és szükség lehet a szövegen belüli keresésre. Az adatok mérete tehát nem mindig határozható meg előre, ezért a változó mezőhosszúságok kezelésére van szükség.

¹⁴ Az egységesített besorolási adatok állománya valójában tezausz, de nem származtatott, hanem az eredeti ismérveken alapul. A dokumentumleírásokban szereplő, azonos dologra (pl. ugyanarra a szerzőre, közreadóra stb.) utaló eredeti ismérveket egy egységesített ismérvhöz rendelik hozzá (pl. az Egyesült Nemzetek Szervezete, ENSZ, UN, UNO nevekkel jelölt közreadót egységesen az ENSZ megnevezés képviseli). Akármelyik nem egységesített kifejezés alapján keresnek, mindegyik célinformációt megkapjuk, amelyet valamelyik helyettesített ismérv jellemez. E kapcsolat a tezauszok nem deszkriptor – deszkriptor relációjának felel meg.

A tezauszokkal és osztályozási rendszerekkel ellentétben erről az állományról hagyományosan nem készül külön szótár, hanem csak magában a katalógusban vannak jelen az egységesített adatok. Gépi körülmények között azonban nem lehet akadálya annak, hogy csak erről az állományról külön szótárszerű kinyomtatás készüljön, feltüntetve a kapcsolatokat is, és akár az egyes lexikai egységekhez tartozó tételek számát. Egy ilyen szótárszerű kinyomtatás megkönnyíti az áttekintést és a karbantartást.

Elsősorban a könyvtári (bibliográfiai) és szakirodalmi célinformációkat jellemzi rendkívül heterogén és nagy felhasználói kör. Felhasználásuk lényegében nyilvános. E célinformációkat a lehető legtömörebb, helytakarékos, sorfolytonos formában jelenítik meg, mivel dokumentumokban is szerepelhetnek (például szakirodalmi hivatkozások formájában), vagy összességük alkot dokumentumot (például bibliográfiák, növekményjegyzékek, referáló kiadványok formájában). Ezzel szemben a múzeumi, levéltári, egészségügyi (lelet) stb. nyilvántartások "dokumentum" célinformációit többnyire úrlapszerű, táblázatos formában jelenítik meg. A nyilvános felhasználás következtében a szakirodalmi és könyvtári "dokumentum" célinformációkat kezelő rendszerek fejlettebb, a természetes nyelvhasználat-hoz közel álló nyelvi komfortot és viszonylag könnyű megtanulhatóságot igényelnek. Ez különösen a mikroszámítógépeken használható adatbázis-kezelő rendszerekre vonatkozik, melyek üzemeltetéséhez eleve nem lehet alkalmazói programozókra számítani.

4.7 A dokumentumleírás, illetve részletei mint célinformációk

A dokumentumot vagy a dokumentumként kezelt tárgyat, eseményt stb. azonosító információk összessége a dokumentumleírás (a 4.4 fejezetben említett "elérési formátum" gyakorlatilag ezzel azonos). A dokumentumleírásnak nagyon különböző fajtái lehetnek. A bibliográfiai leírások a könyvtári gyakorlatban használt szabványos dokumentumleírások (lásd a melléklet bibliográfiai szabványait), a hivatkozási leírások a közleményekben használt dokumentumleírások stb. és a dokumentumleírás nem minden lehetséges elemét tartalmazák (pl. a bibliográfiai leírásnak a szabványok szerint nem része az osztályozási jelzet, a raktári szám vagy a referátum szövege).

A "dokumentum" célinformáció túlnyomórészt a dokumentumleírás szokott lenni. Egyáltalán nem ritkák azonban azok a célinformációk, melyek a dokumentumleírás egy-egy részletéből állnak csupán. Például, amikor arról tájékozódunk, milyen gyógyszer kapott a beteg (azaz nem az egész leletet keressük), mennyibe kerül egy múzeumi tárgy, mi a telefonszáma az egyik olvasónak, egy adott témával milyen szerzők foglalkoztak, egy adott évben milyen magyarországi kiadók jelentettek meg könyveket német nyelven stb., vagy éppen mi a tárolt dokumentumok átlagos mérete, ára.

Különösen jellegzetes dokumentációs eset, amikor ismerni kívánják mindazokat a szavakat,

melyekkel a célinformációkat például tartalmilag jellemezték (de szóba jöhet az összes szerző, kiadó, betegség, foglalkozásnév, növényvédőszer-hatóanyag stb. is). Erre egyrészt azért van szükség, hogy lássák, mit használtak fel eddig a keresés tárgyainak jellemzésére (ígyen javítva a jellemzés vagy a keresés munkáját), másrészt azért, hogy a szókészletet nyelvileg egységesítsék, a szinonimákat kiszűrjék, a szavak között meghatározzák a fontosabb szemantikai kapcsolatot (tezauszust készítsenek), a szerzőneveket stb. esetén kialakítsák az egységes besorolási adatok állományát (az "authority fájl") stb., magyarán: hogy szótárakat készítsenek.

Ilyen célinformációk természetesen csak azok lehetnek, melyek egyben a keresés tárgyainak is jellemzői, melyek alapján tehát keresni lehet. Ha ezek a jellemzők maguk lesznek a keresés tárgyai, akkor szótári célinformációkról beszélünk.

4.8 Az ismérv (kulcs)

A keresés tárgyainak jellemzői közül kerülnek ki az ismérvek; ezek biztosítják, hogy a célinformációk megtalálhatók legyenek. Az ismérv mindig része a célinformációnak, noha önmaga is lehet célinformáció (ekkor vagy egy másik ismérv alapján található meg, vagy teljes állományukat jelenítik meg, mint a szótári célinformációk esetén általában). Funkcionális szempontból három fajtájuk van:

- ◆ Az azonosító ismérvek a mindenkori tételt egyértelműen meghatározzák. Egy azonosítóhoz mindig csak egyetlen tétel tartozhat (azaz értékeik a tételek vonatkozásában egyediek). A számítástechnikában elsődleges kulcsoknak nevezik őket (*Halassy, 1978, p. 428.*). Az azonosítók esetében természetesen mindig pontosan meg kell határozni, mit értünk tételen. Az ISBN- vagy az ISSN-szám például a kiadványtételt azonosítja (a dokumentumot mint kiadási, terjesztési egységet). A leltári és a raktári szám (lelőhely-azonosító) elvileg a birtokba vett dokumentum fizikai egységeit azonosítja, de e téren éppen a könyvtárakban sok következtelenség is tapasztalható. A leltári szám olykor csak kiadványszinten azonosítja egyértelműen a dokumentumot (vagy valamiféle köztes szerepet tölt be a kiadvány és a fizikai szint között), a raktári szám pedig a tétel tárolási helyét azonosítja csak egyértelműen (ezért is nevezik többnyire lelőhely-azonosítónak). A gépi hordozón rögzített tételek esetében rekordazonosítóról, azonosítószámról beszélnek, melyet többnyire a bevitelkor, a beviteli sorrend szerint rendelnek az információtételhez. Nem kell azonban feltétlenül számnak

lennie, lehet összetett adat is. Lényege: mindig csak a kezelt állományban szereplő tételt azonosítja és semmi mást.

- ◆ Az ismérvek másik része a mindenkori tételek közül egynél többet is jellemezhet (azaz értékeik a tételek vonatkozásában nem egyediek); a számítástechnikában másodlagos kulcs a nevük. Egyik fajtájuk a dokumentumból közvetlenül megállapítható, abban – vagy tárgyak esetén azon – eredetileg is szereplő jellemzők (szerzők, cím, ISSN-szám, lapszám, lakcím, hajszín, márkanév, gyógyszer-hatóanyag stb.) vagy tulajdonságok (pl. a tárgy mérete, anyaga, feliratai). Mivel a dokumentumban többnyire megvannak, a tárgyon szemmel láthatóan észlelhetők, ezért formálisan, az alaki (nem tartalmi, nem tematikus) megismerés alapján állapíthatók meg. A dokumentum bevitelkor minden azonosító ismérv, amely addig a dokumentumot valamilyen értelemben meghatározta (pl. az ISBN-szám, a raktári szám), a kezelés szempontjából elvileg az utóbbi ismérvkategóriába tartozik (az információtétel egyértelmű azonosítására nem feltétlenül használhatók).

- ◆ Ugyancsak a másodlagos kulcs szerepét játszó ismérvekhez tartoznak az eredeti dokumentum tartalmát, rejtett tulajdonságait leíró hosszabb szöveges vagy sűrített, jelzésszerű adatok, melyek többnyire csak a dokumentum tartalmának, témájának, eredeti rendeltetésének, a vele megtörtént eseményeknek stb. ismeretében állapíthatók meg, azon eredetileg, legalábbis a kívánt formában nem szerepelnek.

A formálisan megállapítható ismérvek többsége eredeti információ. Ilyenek például a bibliográfiai, az irattári, a leltári adatok túlnyomó része, a leletnyilvántartásban a beteg személyi adatai, a bélyegkatalógusban a bérmentesítés értéke, a régészeti nyilvántartásban a tárgy lelőhelyének neve, egy múzeumi nyilvántartásban a festmény festőjének neve, a kép mérete stb. Természetesen előfordulhat, hogy az eredeti ismérv hiányzik a dokumentumról; megállapítása akkor sem jelenti azt, hogy ezáltal a dokumentum tematikáját, szemantikai tartalmát vagy értelmét, jelentőségét, szerepét jellemezték volna.

A dokumentum mélyrehatóbb ismeretében megállapítható, túlnyomórészt utólag feltárható ismérveket származtatott ismérvek: a tárgyszavak, deskriptorok, osztályozási jelzetek, a tartalmi kivonatok – az annotációk, referátumok – szavai is. Ez utóbbiak, akár csak a cím szavai, a szabad, szövegen belül végzett kereséskor ("free text searching") jätsszól az ismérv szerepét, de kreshetők az ún. csonkolással ("lebegő kereséssel", "string-kereséssel") is. Ma már készülnek mikroszámítógépekhez

is teljes szövegben végzett keresésre alkalmas adatbázis-kezelő rendszerek (McNamara, 1985, FIRS, 1981), de a keresésnek ez a módja ma még nem igazán alkalmazási területe a mikroszámítógépeknek.

Eredeti és származtatott ismérvek helyett formai és tartalmi ismérvekről is beszélhetünk. Nem minden eredeti ismérv formai, ahogy nem minden származtatott tartalmi. A dokumentumról sokszor hiányzik a szerző, s csak komoly feltáró munkával állapítható meg a kiléte; a dokumentum címe mindig eredeti ismérv, de ugyanakkor tartalmi; az osztályozási jelzet vagy a tárgyszó eleve szerepel a dokumentumban, holott szintén tartalmi; az oldal-számot olykor nem tüntetik fel, tehát származtatott; megfelelően tájékoztató cím hiányában a kiadó vagy a közreadó is némi felvilágosítást nyújthat a tartalomról stb.

Az eredeti és származtatott, illetve a formai és tartalmi ismérvek tehát egymást átfedhetik (de az eredeti és a származtatott egymást ugyanúgy kizárja, akár a formai a tartalmi). Az adatbázis-kezelő rendszerekben azonban az egyes ismérvfajtákat többnyire különbözőképpen kell kezelni, mert rendkívül eltérő elvárások fűződnek hozzájuk mind a feldolgozók, mind a felhasználók részéről. (A szerződési adatok feltüntetése és az azok szerint végzett besorolás iránt a bibliográfiai rendszerekben például speciális kezelési követelmények lépnek fel, melyek teljesen eltérnek például a tartalmi ismérvek iránt támasztott igényektől.)

Az ismérvek (a kulcsok) nem azonosak a keresőszavakkal és egyéb, a kereséskor használatos elemekkel. Nem törvényszerű, hogy minden keresőelemnek megfeleljen a tárolóban ismérv és fordítva.

4.9 A célinformáció-tételek megjelenítése

A "dokumentum" célinformáció-fajták esetében szükség lehet mind a teljes, mind a részleges dokumentumleírás megjelenítésére. Bibliográfiai igényű tájékoztatás esetén a leírás szabványos formájára van szükség, egyszerű növekményjegyzékek esetében a mindenkori felhasználói kör igényei szerint rövidítendő a leírások. Közlemények hivatkozási tételeinek megjelenítése céljából a lehető legrövidebb leírást használják, s az is előfordul, hogy találatként elég, ha csak a cím vagy a szerző, vagy a kiadó, vagy a lelőhely, vagy a gyógyszer neve jelenik meg stb.

De nemcsak a teljes vagy részleges leírás között kell tudni válogatni a kezelőrendszer segítségével, hanem a leíráson belüli ismérvek (bibliográfiai szempontból: adatelemek) sorrendje is esetről

esetre változhat. Bibliográfiai igények esetén például előbb a cím, aztán a szerző, a hétköznapi használatban gyakoribb, ha előbb a szerző következik, de bizonyos szolgáltatások esetén arra is szükség lehet, hogy a referátum szövege kerüljön a dokumentumleírás élére, és csak utána következzen néhány elengedhetetlen azonosító adat.

Könyvtári és szaktájékoztatói körülmények között általában a tételmegjelenítés formája sorfolytonos. Az egyes ismérveket (bibliográfiai adatelemeket) szabványosított jelekkel választják el egymástól. Amint más sorrendet kívánnak az ismérvek között, más elválasztó jeleket is kell tudni meghatározni a kezelőrendszerrel. (Ha például a szabvány szerint előbb a cím következik, akkor az utána következő szerzőt a köz, ferde törtvonal és köz választja el a címtől; ha fordítva, a szerző után általában a kettőspont következik.)

A 8. ábrán az egyik legbonyolultabb "dokumentum" célinformáció-fajta, egy időszaki kiadvány — és ezen belül egy folyóirat, a "a Mai magazin" című havilap — dokumentumleírása látható.¹⁵ Alatta feltüntettük a tétel funkcionális-logikai kapcsolatainak modelljét is. E kapcsolatok a tételen belül a kapcsolódó tételek azonosítóinak (pl. → 76/176 vagy →B1) formájában vannak jelen.

Az állományi egységek (az egyes folyóiratszámok) közül csak egyet — a 14. évfolyamot — tüntettük fel, a hozzájuk kapcsolódó könyvtári lelőhelyazonosítókkal (azaz a B1 jelű könyvtárban a 14. évfolyam minden egyes száma megtalálható stb.). A rövid nyilacskák éppen azt emelik ki, hogy tételek közötti funkcionális kapcsolatról van szó (tartalmuk a "kapcsolódó tétel" mezőnévnek felel meg).

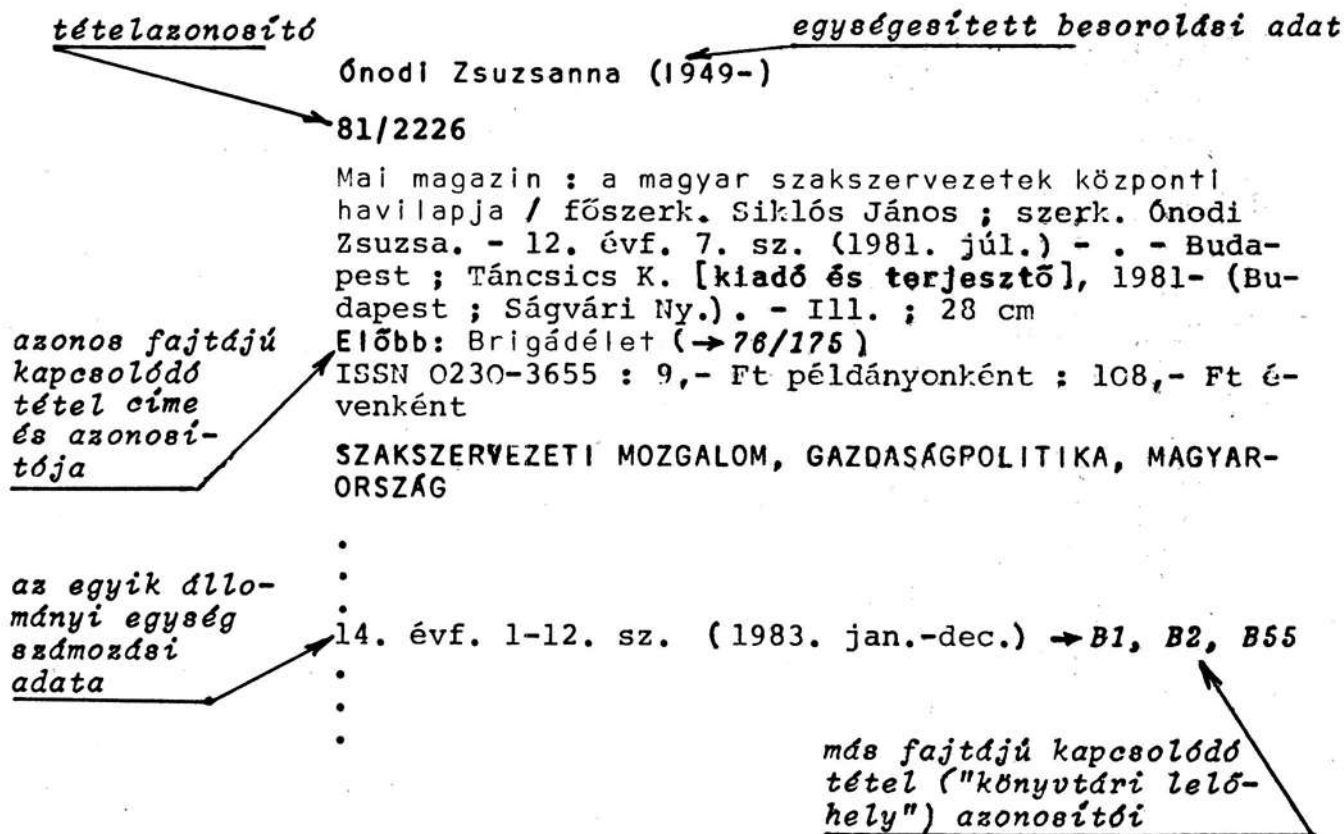
Az ábrán látható dokumentumleírás egésze, valamint kijelölt ismérvei (szabványos adatelemei) tényinformációk. A tényinformációkon — például a címen, a nagybetűkkel írt deskriptorokon — belül minden információ szöveges (így például a "magazin" vagy a "mozgalom" szavak).

Az elsődleges ismérveket vékonyan, a származtatott ismérveket vastagon írtuk. A formai ismérveket talpas betűkkel, a tartalmi ismérveket talp nélküli betűkkel írtuk.

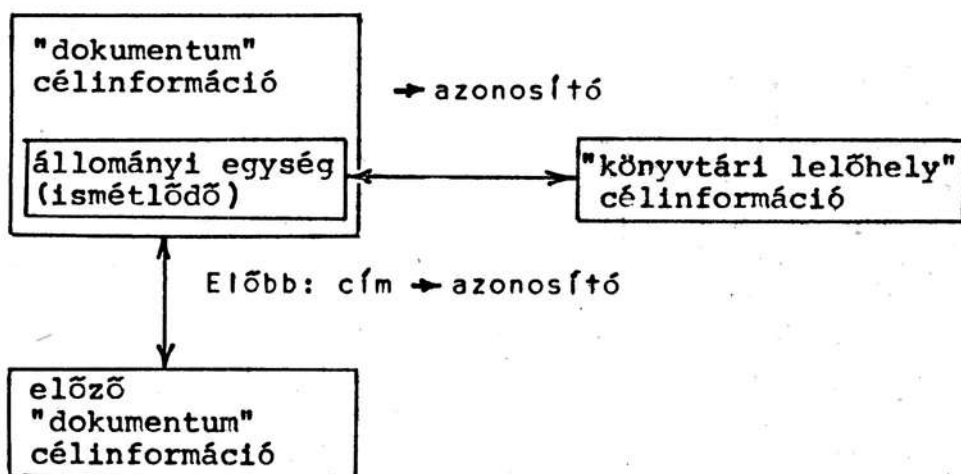
A 81/2226 tételazonosító az adott kiadványban egyértelműen azonosítja a tételt: nem biztos, hogy a tárolt állományon belül ugyanez az azonosító érvényes. Magának a folyóirat fizikai előfordulásának

¹⁵ A tételt — a szemléltetés kedvéért némi kiegészítéssel — a "Magyar Nemzeti Bibliográfia. Időszaki kiadványok bibliográfiája" című kiadványból vettük. Dokumentációs adatbázis-kezelő rendszerek megjelenítési képességeinek tervezésekor — még nem könyvtári esetekben is — célszerű az ilyen kiadványokat tanulmányozni.

FOLYÓIRAT CÉLINFORMÁCIÓ-ISMÉRVEI



A CÉLINFORMÁCIÓ FUNKCIONÁLIS FÜGGŐSÉGI VISZONYAINAK MODELLJE:



8. ábra Folyóirat célinformáció-ismérvfajtái és funkcionális kapcsolatai (magyarázatát lásd a szövegben)

mondjuk a BI könyvtárban szintén teljesen más azonosítója lehet. A dokumentumleíráson belül szereplő ISSN azonosító ismérv pedig sem a fizikai tételt, sem az itt látható dokumentumleírást nem azonosítja egyértelműen, hanem csak a folyóiratkiadvány létét, azt, hogy ilyen folyóiratkiadvány van. Látható tehát, hogy az azonosító ismérvek (az elsődleges kulcsok) mindig csak meghatározott szempontból, meghatározott szinten töltik be egyedi azonosító feladatukat.

Múzeumi, irattári, műszaki, egészségügyi, bélyeg-, anyag-, tárgy- és egyéb nyilvántartási körülmények között gyakoribb a táblázatos, űrlapszerű megjelenítési formátum. Szabványosításuk korántsem annyira előrehaladott, mint a könyvtári-bibliográfiai formátumoké (lásd a mellékletet). Az egyes ismérveket (rovattartalmakat) többnyire természetes nyelvű, első olvasásra érthető rovatnevek (mezőnevek) azonosítják, megadva az ismérv fajtájának, típusának a nevét (pl. "Ára", "Tünetek", "Lelőhely" stb.). A táblázatos kimutatások áttekinthetőbbek, mint a sorfolytonos megjelenítések, különösen jó szolgálatokat tesznek változó adatok nyilvántartásakor, ezért kizárólagosak az irányítási-ügyviteli rendszerekben. Dokumentációs körülmények között a két megjelenítési formátumfajta akár keveredhet is. A katalógusokkal kapcsolatos szerkesztési munkákban például előnyösebb lehet a táblázatos formátum, a találatok kiadásához viszont a sorfolytonos. Ezért dokumentációs adatbázisok kezelőrendszereiben a megjelenítési formátumnak választhatónak kell lennie.

4.10 A célinformáció-állományok megjelenítése

A tárolt célinformációk teljes vagy részleges, de a felhasználás érdekében rendezett és megjelenített állományait "dokumentum" célinformációk esetében (elsősorban a könyvtárakban) katalógusnak nevezik. Elsősorban az eredeti dokumentum megtalálását teszik lehetővé. Ha az állományok áttekintő ismertetése, az állomány egészéről vagy részletéről adódó tájékoztatás a cél, könyvtári gyakorlatban bibliográfiáról vagy növekményjegyzékekről stb. beszélnek. Értelemszerűen vonatkozik ez a filmdokumentumokra, amikor a filmográfiáról beszélnek stb.

A katalógus kifejezést tárgyak, áruk, kereskedelmi forgalomba kerülő tételek esetében is használják. Szerepük ugyancsak az eredeti dolgok beszerzésének biztosítása (csapágkatalógusok, áruházi katalógusok, bélyegkatalógusok stb.).

Ha az állomány megjelenítése csak a tárolással foglalkozó belső használatra, a tételek azonosítására és meglétének ellenőrzésére szolgál, a célinformációk megjelenített állományát leltárnak nevezik.

Az igazgatásban, ügyvitelben a megjelenített állományokat nyilvántartásnak hívják. Elvileg egy katalógus is nyilvántartás, csak éppen olyan, melynek szerepe a szélesebb felhasználói kör tájékoztatása.

A funkcionális kapcsolatban levő irányítási-ügyviteli célinformációk állományát általában táblászerű, strukturált formában jelenítik meg.

A jegyzékek formájában megjelenített állományokon belül a tételek rendezése iránt rendkívül változatos igények jelentkeznek még egy és ugyanazon felhasználói kör esetén is. Könyvtári körülmények között például a katalógusokat a szabványos besorolási adatok állománya szerint kell tudni kezelni, adott esetben kiiktatva a névelők stb. szerepét a rendezésben, anélkül, hogy e névelők például a címből, a dokumentumleírásból eltűnnének. Különleges követelmények lépnek fel az egynél több szerzőt tartalmazó bibliográfiai tételek esetén, akkor, ha nincsen szerző stb.¹⁶

A megjelenítendő állományokat vagy a kereséskor talált tételeket nagyon különböző ismérvfajta szerint kell tudni rendezni. Mivel a keresést gyakran a külső felhasználók — például maguk az olvasók is — végezhetik, a sokféle megjelenítési lehetőségre vonatkozó utasításoknak a lehető legegyszerűbbeknek kell lenniük, nem igényelhetnek semmiféle programozási ismeretet vagy beavatkozást.

A dokumentációs adatbázisok a rendkívül rugalmas keresési lehetőségek mellett éppen a tétel- és állománymegjelenítések változatosságában, kiszámíthatatlanságában különböznek az irányítási célú adatbázisoktól, melyekben az előre rögzített funkciók a megjelenítési igényeket is előre megszabják.

A legkülönbözőbb megjelenítési igények ellenére azonban ugyanolyan marad a dokumentációs adatbázisok kezelőrendszere: továbbra is ugyanarról a kezelőrendszerről van szó.

5. A dokumentációs adatbázis-kezelő rendszer elvi felépítése...

Amikor mindeddig irányítási-ügyviteli, illetve dokumentációs adatbázisokról beszéltünk, a rendeltetés, a funkció szempontjából különböztettük meg őket. Tettük ezt azért, mert a problémák, melyek az adatbázisok és kezelőrendszereik felhasználóit és készítőit érik, e rendeltetések adta keretekben érik őket, és ezért csak e keretek adta nézőpontból van értelme a megfogalmazásuknak. Az elvi működés

¹⁶ A mellékletben közöltük mindazokat a szabványokat, melyek a dokumentációs adatbázis-kezeléskor figyelembe veendőek.

(3. fejezet) és a célinformáció-fajták (4. fejezet) elemzése alapján azonban kiderült, hogy ez utóbbi két szempontból valójában nincs különbség a két-fajta rendszer között.

Szerkezet és elvi működés szempontjából ugyanis a funkcionális-logikai szempontból strukturált, illetve a tartalmi-szemantikai szempontból strukturált adatállományok és kezelőrendszereik között van különbség. Az első esetben beszélnek hagyományosan adatbázis-kezelő rendszerről, a második esetben pedig információkereső rendszerről. Sem a faktográfiai-bibliográfiai, sem pedig a szöveges adatkezelés nem fejezi ki megfelelően azt a kettősséget, amelyről itt szó van. A funkcionális-logikai és a szemantikai-logikai kezelőrendszerek körülbelül olyan elválaszthatatlanok egymástól, akár a nyelv szintaxisa és szemantikája. Ahogy csak a kettő együttese alapján létezik igazán nyelv és beszéd, úgy csak a funkcionális és szemantikai adatkezelés *együttese* esetében lehet szó igazán adatbázis-kezelő rendszerről.

Nem arról van szó, hogy vannak adatok, melyeket funkcionálisan kellene kezelni, és vannak mások, amelyeket szemantikailag, hanem arról, hogy ugyanazokat az adatokat kell tudni mindkét szempontból kezelni egyetlen rendszeren belül.

Az csak a fejlődés korai szakaszából következő, véletlenszerű jelenség, hogy manapság az irányítási-ügyviteli rendszereket szinte kizárólagosan a funkcionális-logikai szempontból strukturált adatbázis-kezelés jellemzi. Valójában szükség volna ezekben a rendszerekben is több-kevesebb tartalmi, szemantikai-logikai – információkereső – kezelésre. Szívesen megfedekeznek arról, hogy minden új technikához kulturális többletre is szükség van. Ezért csak részben igaz, hogy egyelőre a hardver fejletlensége akadályozza meg a szemantikai-logikai adatkezelés elterjedését. A megfelelő látókör és igényesség hiánya is akadályozza.

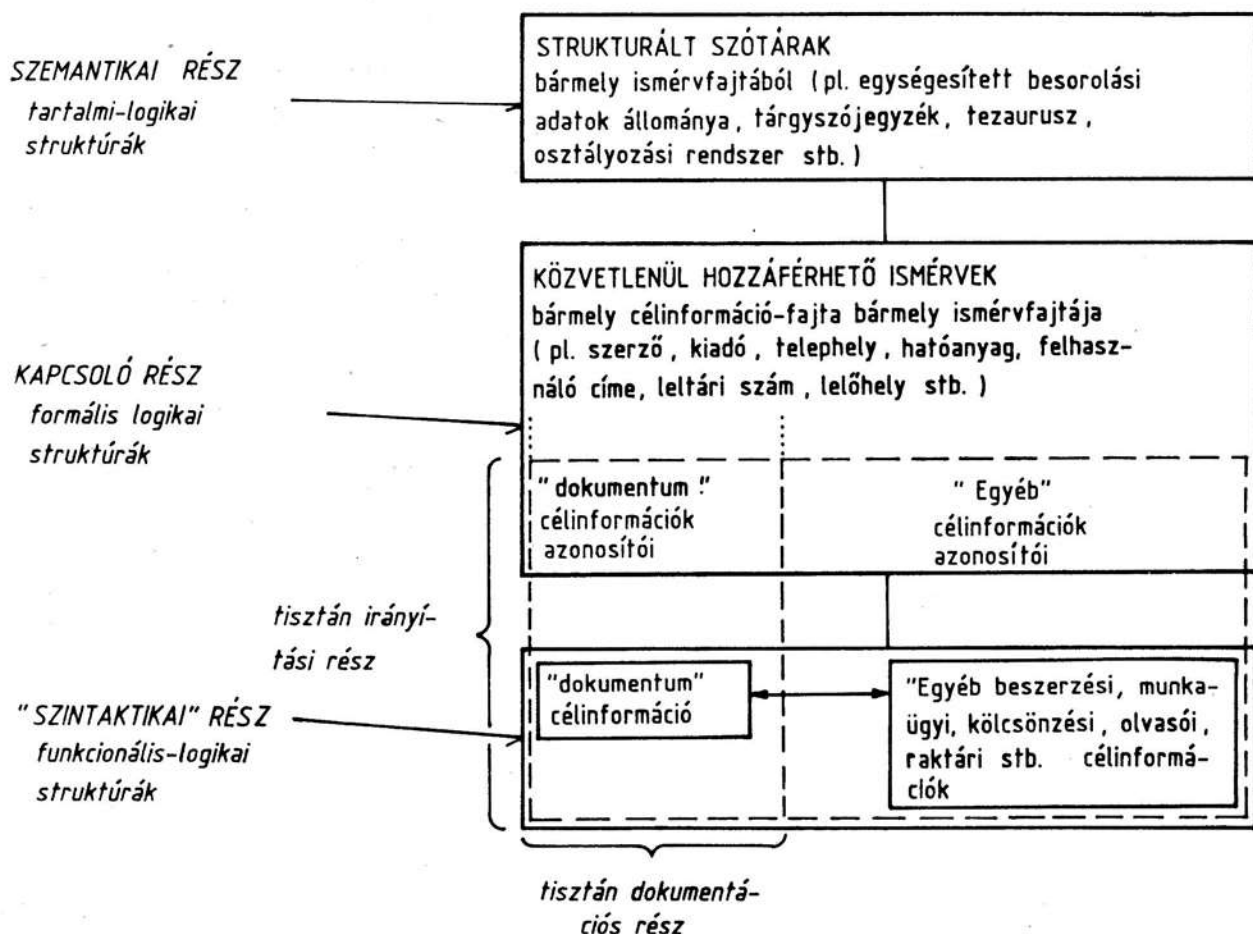
A dokumentációs rendszerekben mindig is létezett több-kevesebb funkcionális vagy adatbázis-kezelő rész. Az állománymegjelenítési igények következtében már a katalógusok kezelőrendszere sem tisztán információkereső. Amiért a dokumentációs rendszerek szemantikai képességei mégis a szükségesnél, fejletlenebbek, annak a hardver adta korlátokon kívül az is az oka, hogy az igazi megrendelők a vállalati-ügyviteli adatbázis-kezelő rendszerek megrendelői, és a számítástechnikai gyakorlat is az ilyen rendszerekhez kötődik manapság. Ennek következtében még túlságosan érvényesül a dokumentáció területén is az irányítási rendszerek mai állapotára jellemző funkcionális programkészítési "filozófia". Jellegzetes példája ennek az a néhány, kereskedelmi forgalomban beszerezhető,

általános mikroszámítógépes adatbázis-kezelő rendszer (például a DBASE, az MDBS – *Armstrong, Weyrauch, Eversberg* közleményeinek tartalmi kivonatait), melyeket lelkesen próbálnak dokumentációs célokra is alkalmazni, holott korántsem sikerültek olyan általánosra, hogy a szemantikai kezelés kérdéseit eredményesen megoldhassák velük. Azért sikeresek mégis, mert a szemantikai kezelés iránti igénytelenség ezt lehetővé teszi. Ez az igénytelenség pedig részben abból ered, hogy a funkcionális kezelés szempontjainak eleve előjogai vannak a gondolkodásban.

Egy dokumentációs célú adatbázis-kezelő rendszer általános logikai modelljét a 9. ábrán foglaltuk össze. A rendszer funkcionális-logikai részének a hagyományos adatbázis-kezelő rendszer felel meg. Ez kezeli az egyes célinformáció-fajták (rekordtípusok) közötti funkcionális összefüggéseket, azaz a "dokumentum", az "olvasó", a "szerzeményezési", a "szelektív információszolgáltatásba bevont", a "felhasználói", a "folyóirat-érkeztetési", a "kötészet" stb. tételfajtákat (a felsorolás természetesen elnagyolt, csak jelzészzerű). Az ábrán a "dokumentum" célinformációhoz funkcionálisan kapcsolódó tételfajtákat az "egyéb" jelzővel foglaltuk össze.

A kezelőrendszer szemantikai-logikai részének a hagyományos információkereső rendszer felel meg. Ez kezeli az egyes – strukturálatlan vagy több-kevesebb strukturált – szótárakat és a szótári egységek közötti szemantikai kapcsolatokat. Elvileg a funkcionális részben szereplő bármely célinformáció-fajta ismerve (adateleme, kulcsa) eleme lehet valamelyik szótárának is, s ezáltal a funkcionális-logikai adatbázisrész minden egyes adateleme bekapcsolható a szemantikai részbe is. A gyakorlatban tárolási megfontolások alapján csak a legszükségesebb szótárakat alakítják ki. Ezek egyrészt a tartalmi, tematikus információkeresést biztosító osztályozási rendszerek, tárgyszójegyzékek vagy teauruszok, a formai ismérvek közül elsősorban a szerzők, testületi szerzők és közreadók nevéből kialakított szabványos besorolási adatok állománya. Az előbbi a gépi szakkatalógus, az utóbbi a gépi betűrendes katalógus alapja. Feltételezhető, hogy a gépi lehetőségek kibontakozásával és a tárolóképeség növekedésével egyéb szótári formák is jobban terjedni fognak az online katalógusokban. Ugyanígy az eddig viszonylag szegényesen "katalogizálható" és "szótárázó" egészségügyi, levéltári, múzeumi stb. dokumentációs rendszerekben is jobban terjedni fog a szemantikai adatbázis-kezelés.

A két rész között egy átmeneti rész helyezkedik el. Tulajdonképpen ez a keresést vezérlő rész. Tartalmazza a közvetlen hozzáféréstől állományokat (az



9. ábra A dokumentációs adatbázis-kezelő rendszer általános szerkezeti modellje

ismérvek és a célinformáció-tételek azonosítóinak kapcsolatait), és itt játszódik le a keresési, összehasonlítási műveletek a különféle logikai operátorokkal (ÉS, VAGY =, ≠, kisebb, nagyobb stb.).

A bemutatott rendszer egésze a tulajdonképpeni adatbázis-kezelő rendszer. Azaz mind a szemantikai-logikai, mind a funkcionális-logikai esetben adatbázis-kezelés játszódik le, csak eltérő módon. Ahogyan nyelvi szabályokat alkalmazunk, akár a mondat szemantikai, akár a szintaktikai részét képezzük. Amint tévedés azt állítani, hogy a szintaxis a grammatika, úgy nem állja meg a helyét az, hogy a funkcionális-logikai kezelés az adatbázis-kezelés.

Egyébként is – a (dokumentációs) adatbázis-kezelő rendszer általánosított modellje nagyon hasonlít arra a modellre, mellyel a nyelv (az állítások) és formális logikája leírható. Ugyanúgy van "szemantikája", (információkereső rész), "szintaxisa" (adatbázis-kezelő rész) és az átmeneti vagy kapcsoló rész formájában "predikátum logikája". Az ismeretbázisú számítógép kialakítására irányuló újabb törekvések valójában ebből a párhuzamból indulnak

ki. Mikroszámítógépen persze még távoli bármiféle ismeretbázis, de a gyakorlatban mégis szem előtt kell tartani ezt a párhuzamot. Többek között abban, hogy csak olyan szoftver készítése vagy vásárlása mellett döntenek, amellyel

- ◆ nemcsak a funkcionális-logikai értelemben vett tételek, hanem minden önállóan kezelhető mezőben rögzített információ és ezek bármely kombinációja is lehet célinformáció;
- ◆ e mezők állományai szótár céljából megjeleníthetők;
- ◆ a megjelenítendő tételek és állományok formátuma szabadon szerkeszthető (mind a sorfolytonos, mind a táblázatos formátumok szerint);
- ◆ változó hosszúságú mezők kezelhetők, és
- ◆ a fenti kezelési műveletek mindegyike alkalmazói programozó beavatkozása, programozási ismeretek nélkül, a felhasználó által használt természetes nyelv alapján legyen megvalósítható.

Ehhez képest az, hogy a kezelőrendszer struktúrája hierarchikus, hálós vagy relációs, a felhasználó szempontjából másodlagos kérdés.

Csak ezek a követelmények állnak némiképp arányban a szoftver szintjén azokkal a lehetőségekkel, amelyeket a hardver fejlődése már eddig is nyújtott, és az optikai lemezek hamarosan bekövetkező térhódításával (Chen, 1985; Price, 1985) nyújtani fog.

6. ...és fajtái

Az elmondottak alapján csoportosítottuk a 10. ábrán a dokumentációs adatbázis-kezelő rendszerek ismertebb fajtáit.

Az egyik végletet a tiszta információkereső rendszerek alkotják: ilyenkor csak "dokumentum" célinformáció-fajta létezik, s ebben csak a tartalom alapján lehet keresni (3. fejezet [g] pontja). Az egyszerűbb változatokban minden szó alapján, amely a funkcionálisan strukturálatlan tételben szerepel, a tétel megtalálható. A fejlettebb változatokban bizonyos szavakat (például névelőket) kizárják a keresésből (ezek a negatív szavak, a nulladeszkriptorok, "stop word", és szótárak a negatív szótár). Még fejlettebb változataikban összeállítják a szótövek és ragok szótárait, és a tárolt tételek szövegében előforduló szavakat velük reprezentálják. A szótövek adott esetben a tezausz deszkriptoraival is összekapcsolhatók (automatikus indexelő rendszerek). Másik változatokban az összetartozási gyakoriságokat elemzik automatikusan a szövegen belül, s ennek alapján képeznek tételjellemzőket, automatikusan kapott ismérveket (ezek az automatikus osztályozó rendszerek). Mikroszámítógépen ma még sem az automatikus indexelés, sem az automatikus osztályozás nem jöhet számításba, e rendszerek egyébként sem kiforrottak még. A szövegen belül szabadon kereső, egyszerűbb rendszerekre már inkább van példa, bár ezek sem gyakoriak. Hozzájuk sorolhatók az indexelő rendszerek: bennük emberi beavatkozással határozzák meg azokat a kifejezéseket, melyek alapján a tétel megtalálható (mutatónevek, indexkifejezések). A szövegben teljesen szabadon kereső rendszerekben a szótár szavait a tételekben szereplő összes szó alkotja; az indexelő rendszerekben csak a kijelölt szavak. Ezek mind tartalmi ismérvek: lehetnek eredetiek, ha a címből, a bibliográfiai leírás egyéb elemeiből vagy az eredeti szövegből származnak, és lehetnek származtatottak, ha a referátum, az annotáció szövegéből származnak.

A katalogizáló, referáló-indexelő szolgáltatásokat kezelő és bibliográfákat, növekményjegyzékeket szolgáltató programrendszerekben ugyan továbbra is csak a "dokumentum" az egyetlen lényeges célinformáció-fajta, de e tételfajtán belül, az egyes tételek között és az egyes tételeken belül is összetett

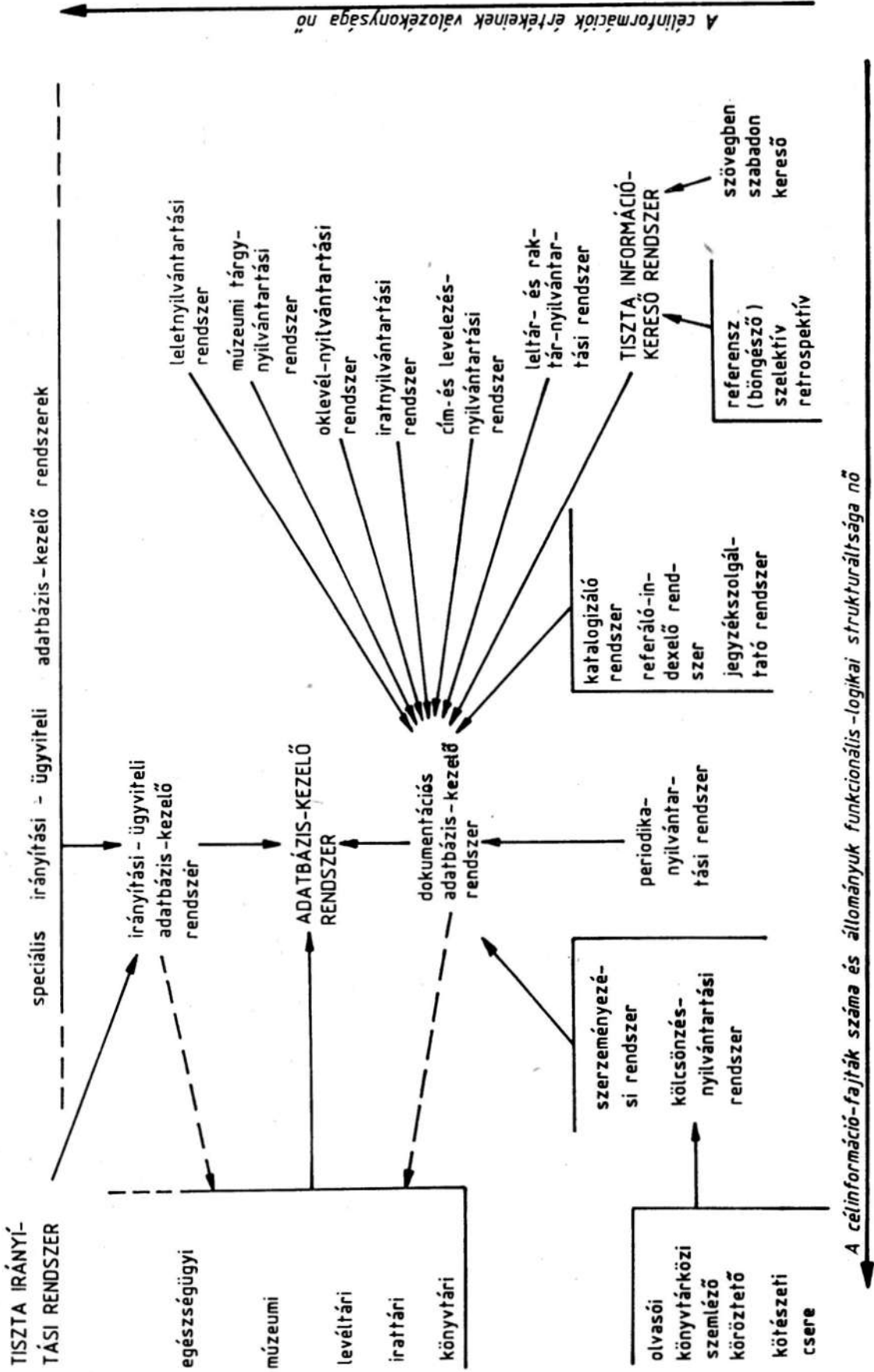
funkcionális logikai kapcsolatokat kell meghatározni. A tételeken belül például előírt sorrendben következnek az ismérvek (adatelemek), az állományokat bizonyos ismérvfajták szerint rendezve kell megjeleníteni.

A katalógusok szerkesztésére könyvtári körülmények között bibliográfiai szabványok vonatkoznak (lásd a mellékletet), a katalógustételek és jegyzék-szolgáltatások tételformátumai iránt akár egyetlen rendszeren belül is változó igények merülhetnek fel (más előírások érvényesek például az egységes besorolási adatok szerint rendezett betűrendes és más előírások az osztályozási jelzetek szerint rendezett szakkatalógusokra). Az állományok megjelenítését rendkívül különböző szempontok – ismérvfajták, mezőtartalmak – szerint kell tudni végrehajtani. Ezen belül egészen különleges esetek is léteznek, mint például a "hivatkozási mutató" ("science citation index"), melyben a "dokumentum" célinformációkat összekapcsolják mindazokkal a "dokumentum" célinformációkkal, melyekre az adott dokumentum szövegében hivatkoznak. A referáló szolgáltatásokban a tételek mindig tartalmazznak hosszabb-rövidebb tartalmi kivonatokat, rendezettségük többlépcsős (például egy olyan osztályozási rendszer szerint, mint az ETO, azon belül pedig például a szerzők, azok hiányában a címek szerint), és mutató is készül az állományokhoz, melyeket permutálhatnak, vagy a mutatónevek (a tárgyszavak, deszkriptorok stb.) között feltüntetnek bizonyos szemantikai kapcsolatokat is "lásd", "lásd még" stb.).

A tartalmi ismérvek ezekben a rendszerekben már többnyire származtatottak, azaz szellemi munkával (osztályozással, deszkriptorok kijelölésével) állapítják meg őket, olykor fejlett szótárakat használva a tartalmi jellemzéshez (például az Egyetemes Tizedes Osztályozási rendszert vagy tezauszot). E rendszerek szótárkezelő képességei – a szemantikai-logikai adatbázis-kezelés – iránt fokozott követelményeket támasztanak.

Ebben a kategóriában kétségtelenül a legbonyolultabbak a folyóiratokat, sorozati és egyéb időszaki kiadványokat (periodikumokat) nyilvántartó, érkezésüket kezelő rendszerek. Itt már összetett funkcionális-állományi kapcsolatok léteznek a "folyóirat" célinformáció és annak részei – az egyes folyóiratszám-tételek – között, azaz vannak "állományi", "bibliográfiai", "érkeztetési" stb. tételek. Egy könyvtár vagy dokumentációs szervezet automatizáltságának színvonala kétségtelenül e nyilvántartás megoldásán mérhető le a leginkább (Holmquist, 1984; McQueen, 1983).

A szerzeményezési és kölcsönzési rendszerekben megjelennek az "olvasói", "beszerzési", "reklamá-



10. ábra A dokumentációs adatbázis-kezelő rendszer fajtái. A relációk (nyílak) magyarázata a 2. ábrához kapcsolódó szövegben szerepel

ciós”, “szállítási”, “raktári nyilvántartási” (vagy “leltári”, “állomány-ellenőrzési” stb.) tételek, azaz a rendszer egyre inkább hasonlít az irányítási-ügyviteli rendszerekhez, mert mind kifejtettebb funkcionális-logikai kezelőrendszere van, ugyanakkor a tartalom szerinti keresés kevésbé fontos.

Egy másik irányban elindulva a tisztán információkereső rendszertől a nem könyvtári-bibliográfiai, azaz a nem szakirodalmi és irodalmi nyilvántartások kezelőrendszerei rendezhetők. Ezek között a levéltári és irattári rendszerek állnak még a legközelebb a könyvtári-bibliográfiai rendszerekhez, a múzeumi rendszerekben már gyakoriak a jól elkülöníthető különféle célinformáció-fajták (a tárgy “lelőhelye”, “feldolgozója” stb.). Az egészségügyi beteg- és leltáryilvántartás, a különféle anyag- és eszköznyilvántartások átvezetnek raktári készletgazdálkodást kezelő, jellegzetesen irányítási-ügyviteli rendszerekbe. (Ilyen különleges eset például a növényvédő szerek nyilvántartása az engedély okiratok alapján. Itt a “dokumentum” célinformáción belül az egyes adatelemek (mezők) között van összetett kapcsolat, mivel ugyanarra a növényvédő szerre különböző élelmezés-egészségügyi várakozási időket írnak elő az egyes növényfajták esetében stb.).

Az irányítási-ügyviteli adatbázis-kezelő rendszereken belül mind a célinformáció-fajták közötti funkcionális-logikai strukturáltság, mind pedig az információk változékonysága különböző lehet. Elvileg elképzelhető az olyan tisztán irányítási rendszer, melyben csak változó adatok léteznek, és minden egyes információ más és más fajtájú. (Ezt a hiperkomplex rendszert a 10. ábra felső sarkában helyeztük el.)

A dokumentációs – és elsősorban a könyvtári – adatbázis-kezelésben a fejlődés iránya kétségtelenül az integrált rendszerek kialakulása felé tart (Telcs, 1986).¹⁷ Ezekben az információkeresés, a katalógizálás, a szerzeményezés és kölcsönzés, olykor a periodika-nyilvántartás is egyetlen kezelőrendszer keretén belül valósítható meg. A kezelhető “dokumentum” célinformációk száma – az Egyesült Államokban használt mikroszámítógépeken – már elérheti a százezret is. Mindebből a következtetéseket nemcsak a felhasználóknak, nemcsak a programrendszerek készítőinek és adaptálóinak, hanem a beruházóknak is le kell vonniuk. A hazai helyzet mind a gyártott szoftvertermékeket (lásd ugyanebben a számban Ungváry, p. 339–342.), mind pedig a

beruházási gyakorlatot illetően (lásd ugyanebben a számban Szöllösy–Ungváry, p. 214–216.) a felhasználóra nézve előnytelen.

Köszönetnyilvánítás

Ezúton mondok köszönetet Szöllösy Évának, aki e tanulmány készítésének minden szakaszában alkotó módon közreműködött: az egyes kéziratváltozatok átnézéséből származó észrevételei mind e tanulmányt gazdagították, olykor szó szerint beépülve a szövegbe.

Halassy Bélával és Vajda Erikkel számos gondolatébresztő beszélgetést folytathattam a tanulmányban érintett kérdésekről; mindez ugyancsak nyomott hagyott ezen a munkán.

Irodalom

- [1] A dokumentáció, az információ és a tájékoztatás fogalma. Terminológiai vita. – Bp.: OSZK Könyvtártudományi és Módszertani Központ, 1961. 72 p.
- [2] BELING, G.–WERSIG, G.: Zur Typologie von Daten und Informationssystemen. – Pullach bei München: V. Dokumentation, 1973. 144 p.
- [3] Codasyl Systems Committee Technical Report: Feature Analysis of Generalized data base management systems. – May 1971. – IFIF Data Processing Group Stadthouderskade 6 Amsterdam.
- [4] CHEN, Ch.: Micro-based optical videodisc applications. = Microcomputers for Information Management. = 2. köt. 4. sz. 1985. p. 217–240.
- [5] DOMOKOS M-né: Faktográfiai tájékoztatás bibliográfiai módszerrel. = TMT, 30. köt. 1. sz. 1983. p. 11–19.
- [6] FID: Outline of a long-term policy of the International Federation of Documentation. – Den Haag: FID, 1960. 39 p.
- [7] FIRS: A free text information retrieval system for microcomputer. = Program, 16. köt. 1. sz. 1982. p. 39–40.
- [8] FUCHI, K.: A tudás alapú információfeldolgozó rendszerek. = Információ/Elektronika, 1983. 2. sz. p. 70–83.
- [9] GORNOSTAEV, U. M.–KOLOZSVÁRY, S.–ZINOV'EV, S. P.: A felhasználói nyelvi eszközök egységesítése az online információkeresésben. = TMT, 30. köt. 8–9. sz. 1983. p. 342–353.
- [10] Grundlagen der praktischen Information und Dokumentation. Eine Einführung. 2. Auflage. (Laisipen, K.; Lutterbeck, E.; Mayer-Uhlenried, K.H.) – München: K. G. Saur V., 1980. 826 p.
- [11] HALASSY B.: Adatbázisok kezelésének alapvető kérdései. – Bp.: SZÁMOK, 1978. 440 p.
- [12] HALASSY B.: Az információrendszer alapfogalmai. – Bp.: SZÁMALK, 1982. 113 p.
- [13] HALASSY B.: A régészeti adatbázis kialakításának lehetőségei és alapvető feltételei Magyarországon. = Archaeológiai Értesítő, 1986. 1. sz. p. 15–21.
- [14] HOLMQUIST, L.: Periodical management at the Apple Computer Library. = Database, 7. köt. 4. sz. 1984. p. 31–36.

¹⁷ E tanulmányban például 23 darab IBM-kompatibilis mikroszámítógépes integrált könyvtári rendszer szerepel, mely külföldi kereskedelmi forgalomban – túlnyomórészt kulcsra készen – beszerezhető!

- [15] KSH 9014/1982. SK 10. KSH számú közleménye a gépi adatfeldolgozási termékek jegyzékének bevezetéséről. = Hazai szoftver kínálat mikroszámítógépekre, 1984. — Bp.: SKV, 1984. p. 7–9.
- [16] Lexikon des Bibliothekswesens. (Herg. Kunze, H.—Rückl, G.) 2. Auflage. Bd. 1–2. — Leipzig: Bibliographisches Institut, 1974–75. 2112 p.
- [17] McNAMARA, F.: Microsystems software program to offer FINDER and PBS. = OCLC Newsletter, 7. köt. 158. sz. 1985. p. 8.
- [18] McQUEEN, J.—BOSS, R. W.: Serials control in libraries automations options. = Library Technology Reports, 1984. 3–4. sz. p. 90–282.
- [19] PRICE, J. W.: The Library of Congress use of microcomputers in the optical disk pilot program. = Microcomputers for Information Management, 4. köt. 2. sz. 1985. p. 241–250.
- [20] SCHUMACHER, B.: Datenbanksysteme auf Mikro-Computern. = Bibliotheksdienst, 10. köt. 12. sz. 1985. p. 10 421–1054.
- [21] TAGG, R. M.: Bibliographic and commercial databases — contrasting approaches to data management with special reference to DBMS. = Program, 16. köt. 4. sz. 1982. p. 191–199.
- [22] TELCS A.: Szakirodalmi áttekintés a mikroszámítógépek könyvtári alkalmazásáról. Tanulmány. — Bp.: Akadémiai Könyvtár, 1986 (előkészületben).
- [23] UNGVÁRY R.: Application of the thesaurus method to the communication of knowledge. = International Classification, 10. köt. 2. sz. 1983. p. 63–68. (Magyarul: Könyvtári Figyelő, 30. köt. 1. sz. 1984. p. 5–17.)
- [24] UNGVÁRY R.: Ein natürliches System der Gegenstände. — Anwendungen in der Klassifikation I. (Herg. Henzler, R. G.) — Frankfurt/M.: Indeks V., 1985. p. 19–41.
- [25] UNGVÁRY R.: Intensional splitting: An empirical examination of conceptual duality. = International Classification, 13. köt. 2. sz. 1986. p. 7–18.
- [26] WERSIG, G.—NEVELING, U.: Terminology of documentation. — Paris: The Unesco Press, 1976. 274 p.

Melléklet

A dokumentációs információkra vonatkozó magyar szabványok

Adatsere-formátum

MSZ 193/1–83

Tételformátum

MSZ 3424/1–78

MSZ 3424/2–82

MSZ 3424/3–79

MSZ 3424/4–84

MSZ 3424/5–79

MSZ 3424/6–83

MSZ 3424/7–84

MSZ 3424/8–85

MSZ 3437–57

MSZ 3402–80

MSZ 3405–59

MSZ KGST 2011–79

MSZ 3494–67

MSZ 3417–68

MSZ 16105–81

MSZ 16171–60

MSZ 3485/2–79

MSZ–05 KGST 1800–79

Állományformátum

MSZ 3401–81

MSZ 3423/1–83

MSZ 3423/2–84

MSZ 3423/3–85

MSZ 3440/1–83

MSZ 3440/2–79

MSZ 3440/3–83

MSZ 3440/4–85

MSZ 3440/5–79

MSZ 3493–82

Ismérvék

MSZ 3563–80

MSZ 3567–75

MSZ 3574–80

MSZ 3400–77

MSZ 3404–77

MSZ 3410–83

MSZ 3432–85

Mágnesszalagos bibliográfiai adatsere-formátuma. A rekordok szerkezete

Bibliográfiai leírás. Könyvek (Módosítás: 1979)

Bibliográfiai leírás. Időszaki kiadványok

Bibliográfiai leírás. Szabadalmi dokumentumok

Bibliográfiai leírás. Szabványosított kiadványok

Bibliográfiai leírás. Szakfordítások

Bibliográfiai leírás. Disszertációk és szakdolgozatok

Bibliográfiai leírás. Kották

Bibliográfiai leírás. Kutatási és fejlesztési jelentések

Bibliográfiai hivatkozás

Könyvek bibliográfiai adatközlése és belső elrendezése

Folyóiratok fedőlapja, belső elrendezése

A dokumentumok referátumai és annotációi

Ügyviteli nyomtatványok tervezése és szerkesztése. Irányelvek

Postaküldemények feladókönyve

Kézbesítőkönyv nyomtatványa

Iktatás, irattározás nyomtatványai folyószámos iktatáshoz

A gépirás szabályai. A gépelt levél

A gépjáratás-technológiai dokumentáció feliratmezői

A bibliográfiai tételek betűrendbe sorolásának szabályai. (Helyesbítés: 1982)

Leíró katalógusok bibliográfiai tételeinek szerkesztése. Általános előírások

Leíró katalógusok bibliográfiai tételeinek szerkesztése. Könyvek

Leíró katalógusok bibliográfiai tételeinek szerkesztése. Időszaki kiadványok

A bibliográfiai leírás besorolási adatai. Fogalommeghatározások

A bibliográfiai leírás besorolási adatai. Személyek nevei

A bibliográfiai leírás besorolási adatai. Testületek nevei

A bibliográfiai leírás besorolási adatai. Címek

A bibliográfiai tételek besorolási adatai. Földrajzi nevek

A bibliográfiai tételek besorolási szabályai

Könyvek nemzetközi azonosító száma (ISBN)

Időszaki kiadványok és sorozatok nemzetközi azonosító számozása (ISSN)

Térképek nemzetközi azonosító számozása (ISBN)

Nyelvek kódjai

A magyar időszaki kiadványok címrövidítése

Országnevek kódjai

Szavak és szókapcsolatok rövidítése a bibliográfiai leírásban

MSZ 10/5-73	Rajkszámzási rendszerek
MSZ KGST 1363-78	Keltezés számjegyekkel
<i>Szótár</i>	
MSZ 3418-76	Magyar nyelvű információkereső teauruszok szerkezete, részei és formája
MSZ 4000-77	Egyetemes Tizedes Osztályozás, Rövidített kiadás. 1. kötet: Táblázatok 2. kötet: Betűrendes mutató
<i>Karakter</i>	
	A magyar helyesírás szabályai
MSZ KGST 1362-78	A cirill betűk transliterációja latin betűkre
MSZ KGST 359	Az adatfeldolgozás alfanumerikus karaktereinek osztályozása, megnevezése és jelölése
MSZ 3495/1-78	A gépirás szabályai. Jelek
<i>Egyéb</i>	
MSZ 3396-85	Könyvtári-ügyviteli nyomtatványok
MSZ 3406-50	Könyvtári és dokumentációs címfelvételi lap
MSZ 3407-50	Könyvtári és dokumentációs osztólap
MSZ 3448-78	A könyvtári állomány nyilvántartása

**UNGVÁRY Rudolf: A dokumentációs célú adatbázis-kezelő rendszer.
A rendszer helye, elvi működése és felépítése**

Ha dokumentációs, illetve irányítási-ügyviteli adatbázisokról beszélünk, a rendeltetés, a funkció szempontjából különböztetjük meg őket. A megkülönböztetés azért szükséges, mert a felhasználók problémái e rendeltetések adta keretekben merülnek föl. A kétfajta adatbázis kezelőrendszerei alapvetően abban különböznek, hogy az első mindig tartalmaz olyan részt, amely a kezelést a dokumentum tartalmi jellemzői alapján és több szempont szerint végzi. A tartalmi jellemzők részben a dokumentum témáját, belső tartalmát jellemző, többnyire utólag megállapítandó (pl. deskriptor, osztályozási jelzet) ismérvek. E szemantikai-logikai adatbázis-kezelő részt nevezik hagyományosan információkereső rendszernek. A dokumentációs célú rendszerek egy része tiszta információkereső rendszer, de csak ritkán olyan, amely tisztán szöveges adatállományokban keres.

Másik, sokkal jelentősebb részük (pl. a kölcsönzés-nyilvántartó, a folyóirat-nyilvántartó, az integrált rendszerek) funkcionális-logikai részt is tartalmaz, melyben a kezelés alapja elsősorban a célinformáció-fajták (a rekordtípusok) közötti, az adott szervezeten belül meghatározott funkcionális kapcsolat. E rész lényegében egy irányítási-ügyviteli rendszer; hagyományosan csak ezt nevezik adatbázis-kezelő rendszernek.

UNGVÁRY, R.: Database management systems for bibliographic information. Purposes, principles, and system designs.

The distinction between bibliographic and management-administration information databases can be performed by their purpose and functions. The database management systems (DBMS) for these two purposes and functions differ in that the former type contains parts offering database management by content characteristics of documents and, in addition, by several different aspects. These characteristics are either the elements given in the documents originally (e.g., title, author, source), or those attached to the documents in a subsequent phase for describing their content (e.g., descriptors, classification codes). This semantic-logic part of the DBMS is called traditionally the information retrieval system. A fraction of the bibliographic information systems is purely information retrieval systems but only few of them are intended for retrieval from textual databases.

The other, more significant than the former part of DBMSs (e.g., library circulation, serial control, integrated systems) can contain also a functional-logic part the management of which is based on the functional relation between information (record) types within a given organisation. This part forms essentially the management-administration information system which is called traditionally DBMS.

Az általános dokumentációs adatbázis-kezelő rendszer tehát információkereső és funkcionális-logikai adatbázis-kezelő részből áll; ezeket a közvetlen hozzáférést biztosító állományok és rajtuk véggezhető formális-logikai műveletek (a Boole-algebra) kapcsolják össze (átmeneti vagy kapcsoló rész). E hármas szerkezeti felépítés egy szemantikai – formális logikai – szintaktikai elemekből álló struktúrának felel meg. A dokumentációs rendszerek tervezésekor és beszerzésekor e három kezelési feladatot mindig szem előtt kell tartani.

A tanulmány részletesen tárgyalja a dokumentációs adatbázis-kezelő rendszer elvi működését, az általa kezelt információfajtákat, a célinformációkkal szemben támasztott speciális tétel- és állománymegjelenítési követelményeket. Végül a kezelt tételtípusok és az adatfajták, illetve időbeli változékonyságuk alapján csoportosítja a dokumentációs adatbázis-kezelő rendszerek fajtáit.

* * *

УНГВАРИ, Р.: Программная система для обработки документальных баз данных. Место, функционирование и построение системы

Когда речь идет о документальных базах данных, т. е. связанных с управлением и делопроизводством, их различают по назначению и функции. Это различие необходимо потому, что проблемы потребителей возникают как раз в связи с назначением. Системы управления этими базами данных отличаются друг от друга тем, что первая из них всегда содержит такую часть, которая выполняет операции по признакам, характеризующим содержание документов и по различным аспектам. Признаки, характеризующие содержание, включают и исходные данные (напр., название, автор, местонахождение) или тему документа, его внутреннее содержание (напр., дескриптор, индекс классификационной системы). Эту семантико-логическую часть, занимающуюся управлением базой данных, традиционно называют информационно-поисковой системой. Одна часть систем документального назначения является чисто информационно-поисковой системой, но очень редко такой, с которой ведутся поиски в текстовых базах данных.

Другой, более значительный раздел (напр., регистрация абонемента, учет журналов) содержит интегральные системы и функционально-логическую часть, в которой основой управления является, в первую очередь, функ-

The general information DBMS consists of a bibliographic information retrieval part and a functional-logic part connected to each other through files providing direct access and through the formal logic operations (Boolean operations) carried out on them. They form the interconnection or transition part. This threefold structure corresponds to a structure containing semantic-logic-syntactic elements. During the planning or purchasing bibliographic information systems these three management functions should be kept in mind.

The operational principles of information DBMSs as well as the types of information handled by them, and the special file and record structure requirements are discussed in detail. Finally, the types of DBMSs are categorised according to the types of items and data and to their temporal variations.

* * *

UNGVÁRY, R.: Das Dokumentations-Datenverwaltungssystem. Sein Ort unter den Programmsystemen, seine prinzipielle Funktion und sein Aufbau

Aus der Sicht der Funktion und Bestimmung kann unter kommerziellen (oder Steuerungs-) und Dokumentations-Datenverwaltungssysteme unterschieden werden. Diese Unterscheidung ist erforderlich, da sich die Probleme der Benutzer im Rahmen der Bestimmung stellen. Grundsätzlicher Unterschied zwischen den beiden Verwaltungssystemen besteht im Vorhandensein eines Informations-Retrieval-Komponents im ersteren. Damit erfolgt die Verwaltung der Bestände aufgrund der Dokumenteninhalte *und* nach mehreren Gesichtspunkten. Die inhaltlichen Merkmale sind teils im Dokument vorgegeben (z.B. Author, Titel, Fundort), teils charakterisieren sie den thematischen, tieferen Inhalt und müssen meistens nachträglich bestimmt werden (z.B. Deskriptoren, Notationen). Einige der Dokumentations-Datenverwaltungssysteme bestehen ausschliesslich aus diesem semantisch-logischen Teil, aber nur wenige darunter verwalten ausschliesslich volle Texte.

Andere (z.B. die Verleihsysteme, die Zeitschriftendatenbank-Verwaltungssysteme) weisen auch einen funktionell-logischen Teil auf, in dem Grundlage der Verwaltung in erster Linie jene Beziehungen zwischen den einzelnen Datenkategorien

циональная связь между видами целевой информации (типами записи) внутри данной организации. Этот раздел, по существу, является системой управления и делопроизводства; традиционно это называется системой управления базами данных.

Таким образом, общая документальная система управления базами данных состоит из информационно-поисковой и функционально-логической частей. Их связывают массивы, обеспечивающие прямой доступ и проводимые по ним формально-логические операции (Булева алгебра). Это тройное построение состоит из семантического, формально-логического и синтаксического элементов. При планировании и приобретении документальных систем необходимо учитывать эти три задания управления.

В обзоре подробно рассматриваются принцип функционирования системы управления документальной базой данных, виды обрабатываемой информации, специальные требования к целевой информации по записи и воспроизведению массива. Наконец, дается группировка систем управления документальными базами данных по типам записей и данных и их временной характеристике.

(Zielinformationsarten, Rekordtypen) sind, die für die gegebene Organisation, in der das System verwendet wird, bezeichnend sind. Diese Komponente ist im wesentlichen ein kommerzielles Steuerungssystem; traditionell wird sie als Datenbasis-Verwaltungssystem (data base management system) bezeichnet.

Das allgemeine Dokumentations-Datenverwaltungssystem besteht also aus einem Retrieval-System und einem funktionell-logischem Verwaltungssystem, die durch Bestände für den direkten Zugriff und durch formell-logische Operationen (boolesche Algebra) miteinander verbunden werden. Dieser dreiteilige Aufbau entspricht im Wesentlichen einer Struktur aus den Elementen Semantik, formale (prädikative) Logik und Syntax. Bei der Planung und Beschaffung von Dokumentations-Datenverwaltungssystemen ist Qualität der Lösung dieser drei Verwaltungsaufgaben immer vor Augen zu halten.

Behandelt werden die prinzipielle Funktion des Systems, die zu verwaltenden Arten von Informationen (die Datenkategorien) und die Erfordernisse gegenüber der Ausgabe der Informationseinheiten und ihrer Bestände. Zum Abschluss wird eine Typologie der Arten von Dokumentations-Datenverwaltungssystemen aufgrund der Datenkategorien und der Veränderlichkeit bzw. Ständigkeit ihrer Werte gegeben.
