

MTA Számítástechnikai és Automatizálási Kutató Intézet Budapest



MAGYAR TUDOMÁNYOS AKADÉMIA
SZÁMITÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZETE

ADATMODELLEK, ADATBÁZIS-FILOZÓFIÁK

Irták:

HANNÁK LÁSZLÓ

RADÓ PÉTER

Tanulmányok 175/1985

A kiadásért felelős:

Dr. VAMOS TIBOR

Fősztályvezető:

DEMETROVICS JÁNOS

ISBN 963 311 197 8

ISSN 0324-2951

TARTALOMJEGYZÉK

BEVEZETÉS	5
1. A TÜKRÖZÉS ÁLTALÁNOS SÉMÁJA	9
2. INFORMÁCIÓS MODELL ÉS ADATMODELL	13
3. SZÁMITÓGÉPES ADATMODELLEK - A LYUKKÁRTYÁTÓL AZ ADATBÁZISIG	18
4. ADATBÁZIS MODELLEK	24
ÖSSZEFOGLALÁS	31
IRODALOM	33

Data models and the data base philosophy

Abstract

This paper intends to characterize the way in which data models reflect the real world. Some concepts /data model, information model, etc./ are introduced for this purpose, and a model describing the information system development process is presented in this terminology. Characteristic features of some data models - from precomputer ones to the relational - are investigated analyzing their relationship with the socio-logical-technological background.

BEVEZETÉS

Dolgozatunkban komplex információs rendszerek - adatbázisok - kialakulásának, tervezésének néhány kérdését tárgyaljuk. Azt a folyamatot szeretnénk kissé közelebbről szemügyre venni, amelynek során az információ kezelésének módszerei kialakultak és önállósultak; megvizsgálni, hogy a komplex információs rendszerek tervezésében elsődlegesen használt modellek hogyan tükrözik a való világot, mik a fogalomalkotás, az absztrakció kiindulópontjai, és végül, hogyan hat vissza ezeknek a modelleknek a létrejötte a valós folyamatokra.

Az információt tároló és a tárolt információt szelektíven visszakeresni tudó program-rendszereket nevezik *információs rendszerek*nek a számítástechnikában. Az elnevezés a számítógépes adatfeldolgozás megjelenésével terjedt el, és ezért a közvélemény hajlik arra, hogy a számítógépes információs rendszer fogalmát azonosítsa az információs rendszerrel. Bár az tény, hogy az információs rendszerek törvényszerűségeinek feltárása, illetve alkalmazása a nagyteljesítményű számítógépek használatával vált lehetségessé, és ezek elterjedését követően indult ugrásszerű fejlődésnek, a két fogalom mégsem azonos.

A munkamegosztás kezdetleges szintjén is szükség volt az emberek között *információcserére*. Ennek az információcserének a jelrendszereként alakul ki a beszéd, a beszélt nyelv. Az ilyen módon közölt információ azonban elvész - rendszerezésére, utólagos következtetések levonására lehetőség nincs. Az első lényeges lépés a *munka és a munkavégzéshez szükséges információ* szétválásának útján az írásbeliség megjelenése. Az adatok pusztá rögzítése azonban nem elegendő ahhoz, hogy információs rendszerről beszéljünk. A rögzített információ azáltal válik a termelésben és a társadalomban valóban ható erővé, hogy valaki összegyűjti, rendszerezi - és felhasználja. Nyilvánvaló más-

részt, hogy a termelés társadalmi méretű szervezéséhez, az állam irányító és szervező funkcióinak ellátásához szükség van az információ szervezett áramlásának biztosítására, összesítésére. Valóban: tudjuk, hogy az egyiptomi fáraók államszervezetében a termelés irányításához sok év tapasztalatából leszűrt /és tárolt/ információkat használtak fel. Konkrét információs rendszerek leírását találjuk Hammurabi i.e. 2000-ből származó törvénykönyvében. Pontosan rögzíti a kereskedő és annak ügynökei közötti - szigorúan írásos - elszámolási rendszer szabályait. Ennek alapján a kereskedő elkészítette elszámolásait, nyilvántarthatta állandó vásárlóit, figyelhette a kereslet változását, az árak alakulását. Ez pontosan ugyanaz, mint amit mi is elvárunk egy - az adott területen működő - információs rendszertől.

A középkori nagy kereskedőházak /pl. Fuggerek/ komoly információgyűjtő és -rendszerező szervezetet tartottak fenn. Világos volt, hogy az információ haszon és hatalom forrása.

Ezek között az információs rendszerek és a modern rendszerek között persze óriási különbség van, méretükben, bonyolultságukban, működési sebességükben. Az elv és a cél azonban ugyanaz: a valós világ egy darabjának leírása, a lényeges adatok rögzítése, rendszerezése azzal a céllal, hogy az információs rendszert létrehozó szervezet működését hatékonyabbá tegyük.

Ennek a törekvésnek *eszköze* /és nem oka, még csak nem is feltétele/ a számítógép és számítástechnika. Tekintsük át röviden, hogyan hatott az információs rendszerek fejlődésére ennek a vitathatatlanul nagyon fontos eszköznek a megjelenése.

A számítástechnika fejlődésének kezdetét az ENIAC-tól /1946/ az első "igazi" - tehát elektromos - számítógéptől szokás számítani, megemlítve a régebbi mechanikus gépeket /Babbage, Leibnitz, Pascal/. Hatását tekintve azonban a számítástechnika igazából csak az ötvenes évek elején, a tranzisztor és a félvezető dióda ipari méretű alkalmazásával, a gépek nagy tömegben való gyárthatóságával született meg.

Az azóta eltelt három évtizedben a számítógépek látványos karriert futottak be. Korunkban a valóság minden területén

alkalmazzák őket, sőt egyes vulgáris nézetek szerint a korszerűség foka és a használt számítógépkapacitás egyenesen arányos.

Valójában - mint mondtuk - a számítógép csupán segédeszköz, mely hatalmas méretű, megbízható memóriájával, gyors számítási képességével hatékony segítséget képes nyújtani az információ tárolásában és visszakeresésében, valamint a tárolt információból levonható mechanikus következtetések felismerésében. A következtetési készség korunkban - a számítástechnika relatív fejletlensége és a feladat bonyolultsága miatt - igen másodlagos szerepet tölt be, noha vitathatatlanul az "intelligens" rendszerek fejlesztése dominál majd a jövőben /ld. pl. a japán ötödik generációs számítógép projektet/. A fő funkció tehát jelenleg az ügyvitelt, a nyilvántartást segítő információ tárolása és elérése. Hazánkban a gépidő igen nagy hányadát /60-70%/ ilyen célú rendszerek üzemeltetésére /a maradék jelentős részét pedig valószínűleg ezek előállítására/ fordítják, és a trend növekvő tendenciát mutat /ld. [1] 170.old./.

A technológiai lehetőség mellett a *logika, a formális nyelvek és algoritmusok elmélete, az informatika* adott elméleti alapokat a számítástechnika fejlődésének. Ez a lehetőség azonban csak az egyik oldal. Az adatfeldolgozásra orientált számítástechnikai alkalmazások létrejöttében hatalmas szerepe volt annak, hogy az ipari *termelés* folyamatai egyre bonyolultabbá, a termelés volumene egyre nagyobbá vált. A sokszor országokat vagy kontinenseket behálózó, számtalan termék forgalmazásával, illetve előállításával foglalkozó nagyvállalatok irányításához szükséges információ megszerzése a katalógusokra, kartonrendszerekre épülő információfeldolgozó rendszerekkel megoldhatatlanná vált.

A szervezeti egység - az első finanszírozótól, a hadseregtől, egészen a professzionális mikrogépek legújabb felhasználójáig, a kisvállalkozóig - saját apparátusukkal képtelenek voltak az egyre bonyolultabbá váló világban összegyűjteni, rendszerezni, és a szükséges pillanatban elővenni a szükséges információt.

A hatvanas években divatos "információrobbanás" fogalom már a számítógép hatására született meg, de ezt nem lehet ilyen

egyoldaluan felfogni. A számítógép - az eszköz - létrejöttét már megjelenése előtt az "információrobbanás"-ban realizálódott igény sürgette, és ennek az elsősorban társadalmi folyamatnak a gép csak a műszaki feltételeit hozta létre /egyéb eszközök mellett/.

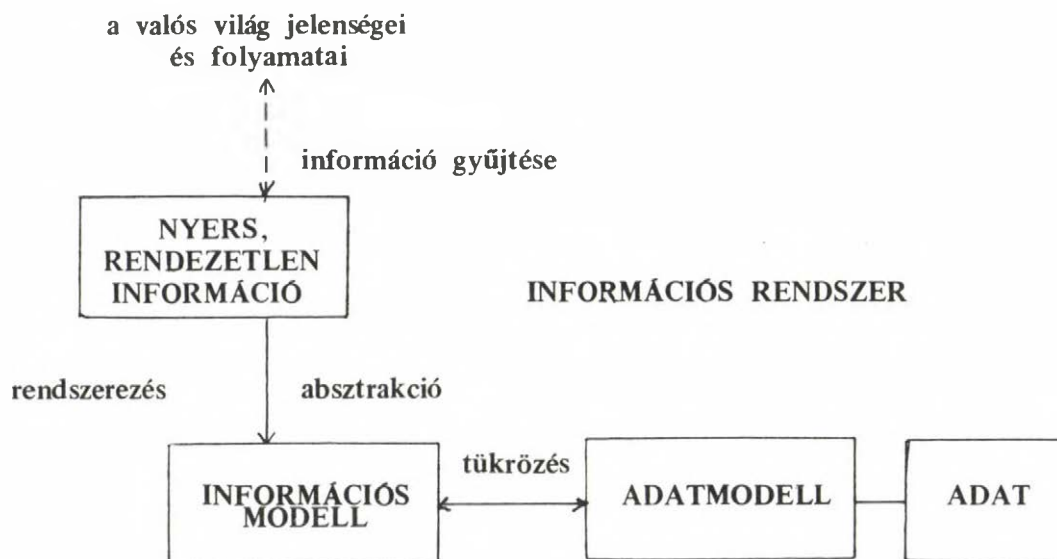
Mindez elég jól látható, ha megvizsgáljuk, hogyan alakult a gépek kihasználása az idő függvényében. Kezdetben a fő felhasználók a mérnökök és a fizikusok voltak, akik a számítások meggyorsítására használták a gépeket /most csak a polgári alkalmazásokat tekintjük/. A számítógép akkor terjedt el igazán, amikor betört a gazdasági életbe, alkalmassá vált üzleti alkalmazásokra. /A piac több mint felét egyedül birtokló IBM ezzel alapozta meg uralmát./ Korunkban a polgári alkalmazások nagy része adatfeldolgozás, sőt a hadsereg is a saját gépkapacitásának jelentős százalékát /pontos számok természetesen nem adhatók/ adatfeldolgozásra fordítja, lévén maga is igen bonyolult belső struktúrával és külső kapcsolatokkal rendelkező szervezet.

Ugyancsak jól illusztrálja tételünket az analóg számítástechnika háttérbe szorulása. A folytonos jelekkel működő igen gyors, de rugalmatlan berendezések a nyilvánvaló hadi és egyéb folyamatirányítási alkalmazások mellett /rakéták célba juttatása, követése stb./ nem találtak alkalmazásra, így jelentőségük a kezdeti lendülethez képes igencsak csökkent.

A folyamat - az információfeldolgozás térhódítása - azonban időt vett igénybe. Ennek - mint az előzőekből nyilvánvaló - nem kizárólag technológiai oka van. Arról van szó ugyanis, hogy a számítógép csupán adatok és utasítások tárolására /és az utóbbiak végrehajtására/ alkalmas eszközként jött létre, de hogy ezt milyen módon célszerű használni /pl. egy vállalat működése miképpen konvertálható át bitekké és programokká/ az igen komoly kutatómunka tárgya volt az 50-es évektől mostanáig, és várhatóan az is marad. A két alapvető összetevő, a hardver és az alapszoftver fizikai és matematikai eszközökkel dolgozó apparátus. Minket a valóságos világot az erre az apparátusra leképző mechanizmus érdekel, melyet nevezhetünk akár adatmodellnek, akár adatbázis-filozófiának /az egyik feltételezi a másikat, és nem is választhatók külön/.

1. A TÜKRÖZÉS ÁLTALÁNOS SÉMÁJA

Kiindulópontunk az 1. ábra lesz, ehhez fűzünk majd megjegyzéseket.



Nyilvánvaló, hogy a *valós világ* / pontosabban annak az információs rendszert létrehozó szervezetet érdeklő darabja / a kiindulási alap. Lényeges megkülönböztetni / mint pl. [2] is hangsúlyozza / magát a valós világot / [2] terminológiájával az anyag-, energiafeldolgozó rendszert /, és az információs rendszert / az 1. ábrán bekarikázva /. Az előbbi maga az anyagi-emberi tevékenység, az utóbbi pedig ebből, és a saját komponenseiből összegyűjtött *információ* rendszerezése, tárolása és feldolgozása. Érdeemes megjegyezni, hogy lévén az utóbbi is anyagi-emberi tevékenység, ebből is gyűjthető információ, és az természetesen tárolható, rendszerezhető. Készíthető tehát meta-információs rendszer, amely az információfeldolgozó rend-

szerek tevékenységéről tartalmaz lekérdezhető információt. Ilyenek az információfeldolgozási tevékenységet leíró szakkönyvek /pl. [3]/, módszertanok /pl. [4]/, és a számítógépes tervező rendszerek mögötti tervezési filozófiák /pl. [5]/.

Az összegyűjtött nyers információból használható *információs modellt* kell készíteni. Az információs rendszer itt használja saját komponenseit /többek között esetleg valamilyen meta-információs rendszert/. Ez alapvetően absztrakciós folyamat, az információhalmaz szerkezetét kell felismerni, a lényegest a lényegtelentől elválasztani. Az így készített modell ideális esetben kezelhető módon választ ad a modellezett szervezettel kapcsolatos valamennyi kérdésre. A gyakorlatban persze a helyzet jóval bonyolultabb, pl. úgy tűnik, semmennyi mutatóval nem lehet leírni, még kevésbé szabályozni egy iparvállalat működését.

Idáig be sem lép a folyamatba a számítógép, és korunkban sem szükségszerű, hogy információs rendszerek kezeléséhez számítógépet használjunk. /Mint már említettük, ezt az információs modell bonyolultsága, vagy a kezelendő információ mennyisége indokolja./ Ez így történt az első információs rendszerektől napjainkig. Hammurabi törvénykönyvének kereskedője az adatokat összegyűjtve összefüggéseket ismerhetett fel, rendszerezhetette /pl. a legtriviálisabb módon területenként/ az adatokat.

A kezdeti információs modellek nyilván igen egyszerűek voltak, a későbbiek egyre bonyolultabbakká váltak, annak megfelelően, ahogy az egyes szervezetek által érzékelhető valóság bonyolódott. [2] a következő okokat sorolja fel az információs rendszerek komplexitását növelő tényezőként:

- *munkamegosztás* /ez nyilvánvaló/,
- *gépesítés* /minden gépet nyilván kell tartani, mert érték/,
- *kapitalizmus* /a tőke teszi lehetővé a nagy vállalkozásokat, viszont a tőkét befektető/k/nek joga /joguk/ van az információhoz. [2] elismeri, hogy ez az informálódási jog még korunkban is sokszor csak látszólagos/,
- *szállítás* /a szervezet tevékenységének szélesíthetősége/,
- *kommunikáció* /nyilvánvaló/.

Tulajdonképpen az információs modellnek a bonyolódása /a minőségi bonyolódás mellett persze az adathalmaz mennyiségének a növekedésére is gondolunk/ vezetett a számítógépes adatfeldolgozás megszületéséhez. Az információfeldolgozási részlegek mérete a babilon kereskedőtől és annak agyagtábláitól pl. a 30-as évek General Motors igazgatóságáig és ügyiratkezeléséig nőtt, és a különféle kézi eszközök száma /különféle elveken épített adattárak, írógép, mechanikus válogatógépek stb./ is gyarapodott. Ennek a folyamatnak logikus fejleménye a számítógép /megalkotói, Norbert Wiener és Neumann János által ilyen mértékben nem sejtett/ "befogása" adatfeldolgozó tevékenységek elvégzésére. Voltaképpen a tükrözés módja a tükrözendő valóság bonyolódása következtében fejlődött és fejlődik tovább napjainkban is.

A számítógép mint adatfeldolgozó eszköz megjelenése előtti eszközök csak kezdetleges *adatmodell* felépítését és használatát engedték meg. Ennek fogalmi a fájl és rekord /amelyeket a mai számítástechnikai nyelvhasználat is örökölt/. Mint később látni fogjuk, a számítógépek bevezetése az adatfeldolgozásba eleinte erre a primitív és "jól bevált" adatsémára épült.

A számítógép megjelenése új adatmodellek bevezetését is lehetővé teszi, azonban az információs modell objektumait le kell képezni a számítógép sokkal pontosabb és ezért sokkal szegényesebb nyelvére. Ily módon jön létre a *számítógépes adatmodell* vagy séma. Ez a számítógéppel tárolt és elérhető adatok szerkezetét írja le. A továbbiakban ennek fejlődéséről és változatairól szólnak majd.

Ez előtt azonban érdemes még egy dolgot megemlíteni. A számítógépes információs rendszer a valóságot tükrözi, annak modellje. Látni kell azonban azt is, hogy létezése vissza is hat a valóságra. Nem csak közvetett módon - nyilvánvaló, hogy a létrehozó szervezet számára az információs rendszer a gyakorlati következményekkel járó döntéshozatalnál létfontosságú, éppen ez indokolja létesítését -, hanem közvetlenül is. Pl. az egyre jobban terjedő hitelkártya rendszer vagy a repülőgépek helyfoglalási rendszere nem képzelhető el számítógépes támogatás nélkül. Az információs rendszer és a termelési folyamat

kölcsönhatásának egy másik vetülete, hogy egy bonyolult termelési folyamatot leíró modell maga is elég bonyolult, és ennek két következményével kell számolni. Egyrészt a modell kialakítása során a termelési folyamatot /nem feltétlenül csak az információszolgáltató részét/ újra át kell gondolni /közben egyszerűsítve vagy ésszerűsítve, vagyis változtatva/. Másrészt a termelési folyamat lokális változtatásának egy *már kialakult* információs rendszer /mind a modell objektív korlátainál, mind az alkalmazók szubjektív korlátozottságánál fogva/ ellenáll.

2. AZ INFORMÁCIÓS MODELL ÉS ADATMODELL

Mielőtt a különféle adatmodellek és a valós világ kapcsolatának vizsgálatába belekezdenénk, szeretnénk pontosan körvonalazni az 1. ábra fogalmait, és néhány lényeges megkülönböztetést tenni.

A "*valós világ*" alatt a valóságnak /természeti jelenségek, társadalmi, termelési folyamatok stb./ azt a részét értjük, amit az információs rendszer létrehozója modellezni kíván. Pl. egy termelőegység munkafolyamatainak modellezésében az időjárás vagy a földrajzi környezet általában elhanyagolható. Azonban speciálisan egy mezőgazdasági létesítmény vagy egy repülőtér információs rendszerében az ezekkel kapcsolatos információ jelentős részére szükség van.

Az "*információ*" fogalma nem tűnik kellőképpen tisztázottnak. Ezt a kifejezést az egyes szaktudományok más és más értelemben használják. A kibernetika és információelmélet ugyan - saját határain belül - pontos definíciót ad, de ez inkább az információmennyiség, illetve egy rendszer strukturális információ-tartalma. A filozófia a strukturális információ mellett /amely a vizsgált rendszer rendezettségét jellemzi/ használja a "relatív információ" fogalmát [6]. Ez mindig két folyamat viszonyával függ össze. Más definíció szerint az információ [7] "valamely eseményre vagy tárgyra vonatkozó, rendszerrel gyűjtött adatok összessége".

Ebben a dolgozatban az információ köznapi értelmezését használjuk, kiemelve két lényeges tulajdonságát.

Általában információ alatt valamely valós folyamatról vagy jelenségről szerzett ismereteink, tudásunk egy részét értjük. Az általunk vizsgált jelenségkörben keletkező információkat az jellemzi, hogy rendezett formában, közlésre alkalmas módon áll-

nak rendelkezésünkre. Az információnak ezt a közlésre, tárolásra alkalmas megjelenési formáját a továbbiakban *adat*nak nevezük. Egy vállalat vezetőjének fejében a vállalat dolgozóiról általában rengeteg /lényeges vagy lényegtelen/ információ halmozódik fel. Adat azonban még az ezek közül lényegesekből is csak akkor lesz, ha a személyzeti nyilvántartásban, bérszámfejtésben stb. rögzítésre kerül.

Az "*információs modell*" a valós világ működéséről, folyamatiról, jelenségeiről összegyűjtött információhalmaz egy absztrakciója. Az információhalmaz lényegtelen részeit elhagyjuk, lényeges részeit kiemeljük, csoportosítjuk, összefüggéseit feltárjuk. A jobb csoportosítás, a szerkezet jobb megvilágítása érdekében új, a valós folyamat jobb leírását segítő fogalmakat vezetünk be.

Az információs modell a valós világnak tehát egy absztrakt képe. Kizárólag az emberi tudatban létezik. Az információs rendszer alkotója /ez lehet több ember vagy szervezet is/ ilyennek látja a valós folyamatot. /Ez az absztrakciós lépés nem feltétlenül történik tudatosan, ami nagyon egyszerű folyamatok esetében elfogadható. Bonyolultabb rendszerek esetén azonban a nem kellőképpen átgondolt, a valóságot torzán tükröző információs modell rossz, pontatlan adatmodell és használhatatlan, a célnak nem megfelelő adatfeldolgozó rendszer kidolgozásához vezet./

Az információs modell fogalomrendszere általában erősen kötődik a valós folyamat fogalmaihoz. A jó információs modell ugyanis elsősorban a valós folyamatra tekint, figyelmen kívül hagyva, hogy a későbbiekben egy adatfeldolgozó rendszer alapjául szolgál.

Valamikor elterjedt gyakorlat volt az információs modell /és ennek alapján természetesen az adatmodell és az ezzel dolgozó programok/ elkészítése a következő módon.

A szervező csupán a modellezendő rendszerbe beérkező adatokat /bizonylatokat/ és az ezek alapján készülő kimutatásokat vette figyelembe, nem törődve a rendszer működésének megértésével, annak pontos leírásával. Maga az adatfeldolgozó rendszer pusztán az így megfogalmazott követelménynek tett

eleget: a beérkező bizonylatokból előállította a kimenő listákat. Ennek a szervezési módszernek a gyengesége az elkészülő rendszer merevségében, labilitásában rejlik. A legapróbb változás /pl. új kimutatás szükséges/ a rendszer egész szerkezetét, a teljes adatmodellt felborithatja. Valójában itt nem is készült el az információs modell, hiszen a rendszernek csupán a külső kapcsolatait irtuk le.

Sok - akár a fentitől eltérő alapgondolatra épülő - kísérlet történt az információs modellek közös tulajdonságainak kiemelése és egy újabb absztrakciós lépés után egy, az információs modellek kidolgozására általánosan használható módszer /a Bevezetés-ben említett meta-információs rendszer/ kialakítására. A teljesség igénye nélkül említünk meg ezek közül néhányat: a "top-down" módszer, a "process-input-output" központu leírások, a TEZAURUS-módszer. Ezek azonban - éppen az információs modellek sokféleségénél fogva eddig csak részleges eredményre vezettek.

Az "adatmodell" az információs modell tükrözése a technikai-technológiai lehetőségeknek megfelelően. Míg az információs modell a valós folyamathoz kötődik, az adatmodell még magasabb absztrakciós szintet képvisel. Az információs modell - noha szükségszerűen egyszerűsítése a valós világnak - még mindig igen bonyolult, adott esetben áttekinthetetlen lehet. Az adatmodell - mely az információs modell realizálása gyakorlati használat céljára - már nem lehet túl összetett, mert akkor nem realizálható.

Az információs rendszereknél megkísérelt általánosítás - mint látni fogjuk - az adatmodellek szintjén sikeresebben tehető meg. Ennek az az oka, hogy az adatmodellek egyszerűbb fogalmakra építenek, és a felhasznált eszközök eleve behatárolják a lehetőségeket. Az adatmodellek körében ez a további absztrakció gyakran terminológiai eredetű félreértésre vezet: a kialakult adatmodell-kategóriákat ismét adatmodellnek nevezik. Mi a továbbiakban adatmodellen mindig egy konkrét információs rendszer információs modelljének megfelelő adatmodellt értünk. Az absztrakcióval keletkező adatmodell-kategóriákat nevezzük /itt és most/ adatkezelési- vagy adatbázis-filozófiá-

nak.

A babiloni kereskedő információs modellje pl. vásárlóinak, ügynökeinek, áruinak, árainak stb. felsorolásából és a közöttük lévő összefüggésekből állhatott. Az adatmodell - az akkori technológiának megfelelően - jóval egyszerűbb volt, a rendelkezésre álló agyagtáblákra jegyezhetette fel egymás után az adatait. Ha elég sok adat gyűlt össze, célszerűnek tűnhetett rendezni őket: egy táblára csak egyféle /egy típusba tartozó/ adatokat írni, és a táblák között is valamilyen rendet tartani.

A számítógép előtti korszak igen bonyolult információs modelleket produkált, gondoljunk csak pl. a General Motors termelési rendszerére, vagy egy-egy minisztérium tevékenységére. Ezek az információs modellek általában a maguk teljességében megfogalmazatlanok maradtak. Ennek egészen egyszerű oka van: a megfelelő adatmodell hiányában ezek az információs modellek realizálhatatlanok voltak.

A *nem számítógépes adatmodell*t lényegében a könyvtárak szervezésével illusztrálhatjuk. Szekvenciálisan /tehát sorosan, adatrekord adatrekord után/ esetleg valami szerint rendezve tárolt adatok - könyvtári katalóguscédulák - és ezekhez indexek - a könyvtárakban pl. könyvcimek szerint. A használatuk nyilvánvaló: ha valaki pontosan tudja, hogy melyik adatrekordra van szükség, és az adatrekord ismert adatára van a tárolás rendezve /könyvtárakban általában szerző szerint/, akkor viszonylat könnyen meg lehet találni az adatot. Ha az adatállomány nincs rendezve az ismert adat szerint, még mindig lehet egy index, amelyik segíthet. Az index maga is rendezett - a nem számítógépes adatmodellben csak a rendezés biztosít gyors elérést az adathoz - és elemei két dolgot tartalmaznak. Az ismert adatot, amely szerint rendezve van /könyvtári példánknál maradvanva ez a könyvcim lehet/, mellette valamilyen mutató /ha a könyvtári cédulák a szerző szerint rendezettek, akkor ez a szerző neve/, amelynek alapján a fő adatállományban a rekord könnyen kikereshető. Ha nincs megfelelő indexe sem, az egyetlen módszer a végigkeresés. Hasonló a helyzet, ha számadatokat kell összegezni: az egész állományt végignézni lassu, fáradtságos munka.

Könnyen látható, hogy az információs modell fejlődése töretlen és egyenletes, hiszen követi a modellezendő valóság változását, a munkamegosztás és a termelési folyamatok egyre bonyolultabbá válását. Egy adott méretű, pl. ipari vállalkozás irányítása és fenntartása ugyan ma sem követel lényegesen több vagy más információt mint 200 évvel ezelőtt, mégis korunkban jóval bonyolultabb információs modellek léteznek, mint akkoriban. Az információs modellek növekvő komplexitása annak a folyamatnak a következménye, hogy maga a modellezendő valóság egyre összetettebb.

Az adatmodellekkel más a helyzet. Fejlettségüket - figyelembe véve, hogy az információs modellek szinte tetszés szerint bonyolultak lehetnek - az adott kor technikai-technológiai színvonala határozza meg.

Az információs rendszeren belül az adatmodell viszonylagos gyengesége jelentette mindig a rendszer legsebezhetőbb pontját. Ez ismét azzal magyarázható, hogy az információs modell a valósággal való közvetlen kapcsolata és az emberi tudat komplexitása miatt jóval bonyolultabb kapcsolatrendszer ábrázolására alkalmas, mint a kor technikájához alkalmazkodni kényszerülő adatmodell. Hadd utaljunk itt Neumann János [8] egy megjegyzésére: a legbonyolultabb számítógéprendszer egyedi építőköveinek és a közöttük levő kapcsolatoknak száma több nagyságrenddel kevesebb, mint az emberi agy idegsejtjeinek, illetve kapcsolatainak száma. A számítógép /és az agyagtábla/ előnye az, hogy nem felejt.

3. SZÁMITÓGÉPES ADATMODELLEK – A LYUKKÁRTYÁTÓL AZ ADATBÁZISIG

A számítógép megjelenése az információs modell komplexitása és az adatmodell viszonylagos gyöngesége közötti ellentmondást nem oldotta fel.

Mint már mondtuk, a számítógépet eredetileg nagy számolásigényű feladatok megoldására hozták létre. /Mérnöki, csillagászati számítások, hadászati ballisztikai táblázatok készítésére./ Mai szemmel ezek a gépek igen lassúak voltak: sebességük 10^2 - 10^3 művelet/másodperc. A tárolható adatmennyiség kezdetben szintén nem haladta meg a 10^3 karakternyi mennyiséget. Egy csillagászati számítás során keletkező adatmennyiség e határon alul van, és lényegében mindegy, hogy az eredmény néhány perc vagy néhány nap alatt születik meg. Adatfeldolgozási feladatokról ugyanezt nem mondhatjuk el.

A másik probléma a megfelelő, olcsó, tartós, megbízható adathordozó hiánya volt. Ez utóbbira hamar született megoldás: a Hollerith-típusú lyukkártyás, mechanikus, rendezőgép által használt papirkártyák alkalmasnak bizonyultak a számítógépes felhasználásra. /Sőt, olyan jól beváltak, hogy sok területen ma sem helyettesíthetők mással./

Ennek a szisztémának nagy előnye, hogy a *lyukkártya* fájlok szerkezete nagyon egyszerű, a gép minden átkódolás nélkül képes volt azokat használni. Ugyanakkor nagy hátrányuk, hogy a felhasználó számára az információ nehezen felismerhető.

A kezdeti adatfeldolgozó rendszereknek ezek a tulajdonságai a mai szempontunkból voltaképpen érdektelenek. Ami döntő jelentőséggel bír az, hogy ezeken az eszközökön lényegében csak ugyanaz az adatmodell realizálható értelmes hatékonysági paraméterekkel, mint ami már a babiloni kereskedő agyagtábláin megvolt. /A lyukkártyás adatfeldolgozás adatmodelljeinél a

könyvtári katalógus adatmodellje /mint modell/ többre használható: a könyvtár karbantartásának információs modelljében a név, cím vagy téma szerinti keresés lényeges momentum, és ezt a katalógus lényegesen hatékonyabban képes realizálni, mint a lyukkártyás feldolgozás./

Vizsgáljuk meg, mi jellemzi azt az adatmodellt, ami pl. lyukkártyás rendszerrel megvalósítható.

A modellben szereplő objektumok és a rajtuk definiált műveletek jelentik tulajdonképpen az adatmodellt. Az objektumok felelnek meg az adathordozón megjelenő adatoknak, a modellben definiált műveletek pedig az adathordozókon-végezhető változtatásoknak. A lyukkártyás adatfeldolgozáshoz tartozó adatmodell objektumai a "fájlok és a rekordok".

Az objektumok és a rajtuk végzett műveletek csak egymáshoz való viszonyukkal definiálhatók. A rekord az adatoknak az a legkisebb egysége, amellyel az adatmodellben definiált műveletek elvégezhetők. Általában egyféle dolog /a modellezendő valóság egy eleme/ egyetlen példányának attribútumaira vonatkozó adatok rendezett összessége.

A művelet az adatmodell objektumain végezhető tevékenység. Általában meg szoktak különböztetni kétféle műveletet:

- keresés: egy adathalmaz kiválasztása,
- módosítás: bizonyos adatok megváltoztatása.

A fájl rekordok névvel azonosított, esetleg rendezett halmaza. A jelen adatmodellben a következő műveletek definiálhatók:

- egy fájl egy rekordjának kiválasztása a teljes fájl átválogatásával;
- egy rekord tartalmának változtatása;
- egy rekord törlése a fájlból;
- egy rekord beillesztése a fájlba.

Most már pontosan meg tudjuk fogalmazni, hogy a könyvtári katalógusokban használt adatmodell mennyivel "tud többet" a fentinél.

Egy könyv kikeresése nem jelenti a könyvtár összes könyvé-

nek végigválogatását. Az adatmodell egy objektummal és több művelettel gazdagabb: az új objektum a "doboz" /partició/, amely rekordoknak egy rendezés szerinti intervalluma, amit az intervallum megadásával azonosítunk, /pl. szerző neve: A-C/. Az új műveletek:

- egy doboz közvetlen kiválasztása;
- egy rekord kiválasztása a doboz tetszőleges helyéről;
- már kiválasztott rekordot közvetlenül megelőző illetve követő rekord kiválasztása.

Az, hogy az adatmodell ezekkel az elemekkel bővithető, azon mulik, hogy a könyvtárban nem számítógépet használunk, mivel az - legalábbis a lyukkártya technológiai színvonalán - képtelen a könyvtáros által rutinszerűen elvégzett tevékenységek ellátására.

A technológiai fejlődés következő lépcsőjét a *mágnesszalag* jelenti. A mágnesszalag - noha technikai megvalósítása jobb - lényegében nem sokban mulja felül a lyukkártya biztosította adatkezelés lehetőségeit. Arról van szó ugyanis, hogy a lyukkártya legfőbb elvi korlátja - az adatrekordokat csak sorban, egymás után lehet elérni - mágnesszalag használata esetén változatlanul érvényes marad.

A hardver következő - a lyukkártyás perifériáktól előnyösen különböző - tulajdonságai okozzák, hogy mégis bonyolultabb adatmodellt használhatunk mágnesszalagos rendszerekben:

- a mágnesszalag több nagyságrenddel gyorsabb;
- a mágnesszalag visszatekerhető;
- a mágnesszalagon megadott számú rekorddal előre vagy hátra léphetünk, a közbülső rekordok beolvasása nélkül.

Ezekből a tulajdonságokból - külön kiemelendő a gyorsaság - adódik, hogy noha a hardver maga nem képes sokkal összetettebb műveletek elvégzésére, mint lyukkártya esetében, mégis megengedhető az adatmodellben a bonyolultabb műveletek definiálása. A feladat ezek után csupán annyi, hogy ezeket le kell fordítani olyan műveletek sorozataivá, melyeket a mágnesszalag egyenként értelmezni tud, és a teljes sorozatot elfogadható időn belül

képes végrehajtani.

Meg szeretnénk jegyezni, hogy a fentebb említett konverzió az adatmodell és a hardver elemi műveletei között a további fejlesztések standard eszközévé vált. A konverziós technika maga is igen látványos fejlődésen ment keresztül, korunk felhasználója és a csupasz számítógép között nagyszámu réteg található. Ezeknek a rétegeknek a feladata, hogy a felhasználó számára egy viszonylag bonyolult adatmodellt biztosítsanak, amelynek lehetőségeit közvetlenül használhatja. Az adatmodell egymásra épülő rétegei egyre közelebb kerülnek a számítógép által használt belső nyelvhez, a legbelső réteg a számítógép működését jelentő elektromos és mágneses állapotváltozások sorozata.

Ez a konverziósorozat - noha főképpen a rendszerek hatékonysága szempontjából igen jelentős - csupán technikai probléma. Arról van szó ugyanis, hogy a döntő kérdést - az információs és az adatmodell viszonyát - nem befolyásolja. Olyan adatmodellek készítésével érdemes foglalkozni, amelyek úgy fordíthatók le a hardver belső nyelvére, hogy a keletkező utasításorozat ésszerű időn belül végrehajtható legyen. A fordítás technikája közömbös.

Már az első számítógépes adatfeldolgozási alkalmazások során megfogalmazódott a következő igény: az adatok halmazából - ez igen nagy lehet - ki kell tudni választani egy meghatározott tulajdonságokkal bíró részhalmazt. Ez adatmodell kérdése: vagy van benne ilyen művelet definiálva, vagy nincs. A gyakorlati tapasztalat szerint nagy adatmennyiség esetén hiába definiálunk ilyen műveletet - noha ennek elvben semmi akadálya nincs - a végrehajtási idő lyukkártyás vagy mágnesszalagos perifériák esetén az alkalmazások szempontjából megengedhetetlenül hosszú lesz.

A probléma ilyen módon való felvetése igen általános. Az információs modellek nagy részének realizációja ezen múlik. Ha képesek lennénk tetszőleges adatmennyiség tetszőleges részhalmazának kiválasztására ésszerű idő alatt, akkor az információs modell adatmodellé transzformálásánál az egyetlen komoly probléma az adatmodell objektumok kijelölése lenne.

Az adatkezelés technológiája szempontjából döntő áttörést jelentett a közvetlen elérésű *mágneslemez tároló* /diszk/ megjelenése. Ez lehetővé tette bármely rekord elérését közvetlen címzés alapján, vagyis anélkül, hogy eléréséhez más rekordok beolvasására is szükség lenne. Ezzel a hardver biztosítani tudja pl. a könyvtári katalógus adatmodelljében szereplő könyvtárosi funkciót: odamenni a megadott dobozhoz, és annak meghatározott helyéről kiemelni egy adott rekordot.

A közvetlen elérés megjelenése lehetővé tette, hogy a hardver szintjén hatékonyan megvalósítható műveletekből bonyolult és változatos adatmodelleket építhessünk. Ezért nem is beszélhetünk egységes, diszkre épülő, akár jó, akár rossz adatmodellről. Az adatmodellek strukturálódását, fejlődését a diszk megjelenése után az eddigieknél sokkal inkább az információs modell minél tökéletesebb megvalósítására irányuló törekvés határozza meg.

A technológiai korlát itt a következő módon jelentkezik: a közvetlen elérés hardver címmel operál. Ennek a modellezett valós világhoz, következésképpen az információs modell fogalmaihoz semmi köze nincs. Ebből adódóan, olyan adatmodellt kellett kialakítani, amelynek objektumaiban keveredtek az információs modellből származó, és a hardver korlátok miatt kényszerűen bevezetett elemek /indextáblák, pointer-technikák, hashing, bitkevesés stb./.

Mindenesetre leszögezhetjük, hogy a fájlkezelésre /vagyis a mágneslemezen a közvetlen címzést biztosító hardverre és szoftverre/ épülő modellek objektumai változatlanul fájlok és rekordok, esetleg ezekből képzett összetettebb szerkezetek /indextábla, kapcsolórekord stb./. Ugyanakkor az ezekben az adatmodellekben szereplő műveletek meglehetősen változatosak, és ez a változatosság ad lehetőséget az egyes információs modellek gazdagságának tükrözésére. Az információs modell elemei közötti kapcsolatok, a folyamatok, ill. a folyamatokat tükröző adatmódosítások leírása azonban esetleges, és az információs rendszer fogalomrendszerétől idegen, a hardver szerkezetéből származtatott objektumokkal történik.

Ez konkrétan a következő problémához vezet: a sokféle fájl

és rekord - ezek jelentős része csupán az adatok közvetlen elérésének az érdekében jön létre - karbantartása nehéz. Ugyanaz az adat - általában megint csupán az elérhetőség érdekében - több helyen is szerepel, ezzel a rendszer redundánssá válik. Ezek látszólag csak technikai problémák, noha már itt feltűnik, hogy a valós világ tükörképe némileg torz, hiszen fájllokba és rekordokba kényszerítettük, redundáns ábrázolást irtunk elő, és ugyancsak technikai okokból olyan objektumokat vezetünk be, amelyek jelenlétét az információs modell nem teszi indokolttá.

A következő alapvető jelentőségű észrevétel vezet el az adatbázis koncepcióhoz: *a cél az, hogy az információs modellnek minél tökéletesebb realizációja legyen az adatmodell.* Az adatmodelltől a hardver belső nyelvéhez vezető leképzéssorozat automatikus. Olyan adatmodellt kell biztosítani, hogy a rendszerszervező célja ne a hardverhez való alkalmazkodás, hanem a minél jobb információs modell megalkotása, és minél pontosabb tükrözése legyen.

Ennek a célnak a fájlkezelő rendszerek a fentebb részletezett okok miatt nem felelnek meg, noha jobbak a mágnesszalagra vagy lyukkártyára építhető adatmodelleknél. A gyakorlati tapasztalatok szerint ezeket használva a rendszerszervező elvész a technikai részletekben, ezért már az információs modell torz lesz, nem odatartozó fogalmakkal terhelt. Ami ennél is súlyosabb: a részletekre koncentráló szervező csak részleges modellt alkot, nem képes átlátni a dolgok valódi szerkezetét, nehezen bővíthető, rugalmatlan rendszert készít.

Olyan adatmodellre van tehát szükség, amely

- elég általános fogalomrendszert ad;
- fogalomrendszere minél közelebb áll az információs modellekéhez;
- nem tartalmaz a technológia esetlegességeit tükröző elemeket;
- még lefordítható a hardver nyelvére.

Ezek voltak az adatbáziskezelő rendszerek és adatbázis-filozófiai megalkotóinak törekvései.

4. ADATBÁZIS MODELLEK

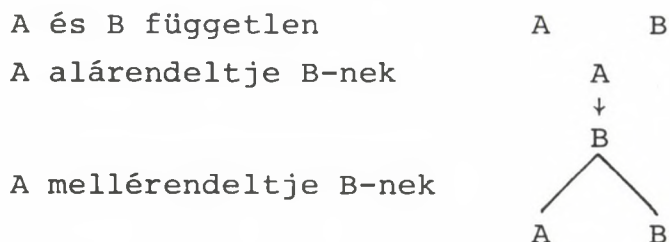
Néhány nagy adatkezelő rendszer megvalósítására irányuló kísérlet látványos kudarca után [9] nyilvánvalóvá vált, hogy az előző fejezet végén felsorolt kritériumoknak eleget tevő adatmodellek a fájlkezelő rendszerek továbbfejlesztésével nem alakíthatók ki. Ugyanakkor az egyes adatmodellek ill. a mögöttük álló információs modellek elemzése rámutatott arra, hogy az egyedi megoldások különböző módon, de hasonló problémával küzdenek. Egyes részfeladatok megoldására /bérügy, raktárnyilvántartás stb./ önálló, de széles felhasználói körben alkalmazható rendszerek születtek. Ezek - a maguk szűk területén is - túl merevnek bizonyultak, elsősorban az eltérő információs modellek miatt. Az a törekvés tehát, hogy a valós világ modellezendő darabjának szűkítésével, az információs modell érvényességi körének korlátozásával nyerjünk jó adatmodelleket, nem vezetett sikerre.

Anélkül, hogy ez a szakmai közvéleményben tudatosodott volna, az információs modellek közös vonásait kezdték el vizsgálni. /Az ún. szoftver-valóság [10] hatására a figyelem a probléma-definiálás, azaz esetünkben az információs modell felé fordult. Nem véletlen, hogy ennek a korszaknak a termékei az olyan meta-információs rendszerek, mint pl. a HIPO, PSL/PSA, SADT és a többi rendszertervező módszertan./

Az információs modellek - és ezen keresztül a való világ - szerkezetének közös vonásait elemezve kiderült, hogy a rendelkezésre álló adatmodellek legnagyobb hiányossága az objektumok közötti viszonyok tükrözésének primitív és esetleges volta. Az az út tűnt leginkább járhatónak, hogy általános sémát keressünk a dolgok közötti kapcsolatok ábrázolására.

A kapcsolatok ábrázolására elsőként alkalmas adatmodell az ún. *hierarchikus modell* volt. A modell hierarchikus volta azt

jelenti, hogy objektumai /változatlanul a rekordok/ egymással alá- vagy mellérendeltségi viszonyban állnak. Ezeket a hierarchiákat az adatmodell műveletei kihasználják. Ha ,A' és ,B' két rekordja a hierarchikus modellnek, a következő lehetőségeink vannak:



Lényeges megszorítás, hogy egy rekordnak csak egy fölérendeltje lehet. Olyan jellegű műveleteket definiálunk az adatmodellben, mint:

- keresd meg ,A' fölérendeltjét,
- keresd meg ,A' összes (első, következő) alárendeltjét.

Ennek az adatmodellnek a műveleteivel mindent meg lehet valósítani, amit a fájlkezelő rendszerek egyedi megoldásaival, de egységes módon.

Az, hogy elsőként éppen a hierarchikus modell jelent meg és terjedt el, alapvetően két okra vezethető vissza. Először is ez viszonylag kevésbé absztrakt modell. Ha körülnézünk a világban, sok, első pillantásra hierarchikusnak tűnő szerkezet, szervezet fedezhető fel.

Az információs rendszerek nagyrésze termelést, ügyintézést, vagy egyéb emberi tevékenységet modellez. Mivel a korszak adatmodelljei a valóság viszonylag szűk területének tükrözésére voltak csak alkalmasak, egyszerűsítő feltevéseket kellett eszközölni már az információs modellben. A fájlkezelő rendszerek alkotói sok esetben folyamodtak ahhoz a módszerhez, hogy csupán egy oldalról ábrázolják a dolgok közötti kapcsolatokat.

Például egy készletgazdálkodási rendszer szempontjából a megrendelések és az áruajták között alárendeltségi viszony van. Egy áruajtának logikailag alárendelhetjük a reá vonat-

kozó megrendeléseket, hiszen a rendszer számára csak az a lényeges, hogy az áruból mennyi fogy el majd egy bizonyos idő alatt. A hierarchikus modell tehát működik.

Az, hogy a rendelésnyilvántartás szempontjából az alárendeltség éppen fordított - itt érdekes, hogy egy megrendelés milyen árukra vonatkozik - a készletgazdálkodást nem zavarja. Mind a két rendszer hierarchikusnak tűnik, és addig, amíg külön-külön funkcionálnak, ez semmi gondot nem okoz. /Más kérdés, hogy a fenti példa kitűnően illusztrálja a fájlkezelő rendszerek redundanciáját./

Látható tehát, hogy mivel a hierarchikus modell a fájlkezelő rendszerek utáni első lépés volt, korlátainak egy részét örökölte. Mégis, ez a nem túl bonyolult séma alkalmasnak látszott arra, hogy általános modellként szolgáljon.

A modell igen szemléletes, számítógépen viszonylag könnyen megvalósítható. Ez népszerűségének másik oka.

A hierarchikus modellel kapcsolatosan meg kell állapítani, hogy noha utólag túlzottan egyszerűsítőnek, primitívnek tűnhet, mégis óriási jelentőséggel bír. Ez az első olyan kísérlet, amely az információs modellek - vélt vagy valódi - közös tulajdonságait az általánosítás egy magasabb szintjén próbálja meg az adatmodellekbe beépíteni, adatbázis-filozófiát teremtve. A rendszerszervező ebben az adatbázis-filozófiában gondolkodva, az információs modell alárendeltségi viszonyait közvetlenül az adatmodell eszközeivel valósítja meg.

A modell gyors elterjedésének kellemetlen következménye volt, hogy az ezt használó szakemberek igyekeztek - az információs modell korrekt meghatározása helyett vagy rovására - ott is hierarchikus kapcsolatokat keresni, ahol a valós világ ennél bonyolultabb volt. A kapcsolatok egyoldalú szemléletével /ld. pl. fenti készletgazdálkodásos példánkat/ ez természetesen megtehető. Ugyanakkor a hierarchikus modell megteremtette a lehetőséget és az igényt nagyobb területet átfogó komplex információs rendszerek megalkotására. Abban a pillanatban azonban, amint megpróbálkoztak az egyes részterületek jól adaptálható információs modelljeinek integrálásával és adatmodellre való leképezésével, kiderült, hogy a hierarchikus szerkezet a kelet-

kező bonyolultabb strukturák modellezésére csak kompromisszumokkal alkalmas. Ugyanaz történt, mint a fájlkezelő rendszerek esetében: redundancia árán, mesterséges objektumok és hierarchiák bevezetésével lehetett a valóság nem-hierarchikus kapcsolatait ábrázolni.

Az alapvető probléma az, hogy az információs modell elemeinek esetleg bonyolult kapcsolatrendszerét a hierarchikus adatbázis-filozófia segítségével egyfajta kapcsolatra kell vetíteni, mert az csak egy szempontu hierarchiát tesz lehetővé.

Nagyon komoly problémát jelent például egy lényegénél fogva hierarchikus szervezet, a hadsereg leírása: a beosztás, illetve rang szerinti alá- és fölérendeltségi viszonyok ugyanis nem szükségképpen azonosak.

A probléma megoldása az adatmodell fejlesztése. Ennek a fejlesztésnek és egyfajta szabványosítási törekvésnek az eredménye a *CODASYL adatbázis-filozófia*.

A CODASYL-filozófia szintén hierarchikus kapcsolatok definiálását engedi meg. Míg azonban a hierarchikus modellben egyetlen hierarchiát kell definiálni, és ezen kívül más kapcsolat a rekordok között nincs, a CODASYL többféle kapcsolatot megenged. Ezt azzal a módszerrel éri el, hogy egyszerre csak két rekordot hozhatunk alá-, illetve fölérendeltségi viszonyba, és egy ilyen kapcsolat létrehozása nem akadályozza meg további hierarchiák létesítését, melyben a két rekord valamelyike /vagy akár mindkettő/ részt vesz. A rekordpárok közötti kapcsolatokat névvel látjuk el, és olyan állításokat mondunk az adatmodell elemeiről, mint hogy:

„A' rekord a „K' kapcsolat szerinti alárendeltje „B' rekordnak.

Az adatmodell műveletei ezeket a kapcsolatokat használják ki. A kapcsolatok megadásával a rekordokon bonyolult kapcsolatgráf adható meg, ezen a gráfon navigálhatunk. A rendelkezésünkre álló műveletek például:

- keresd meg a K kapcsolatban az A rekord első (utolsó, következő) alárendeltjét,
- keresd meg a megtalált rekord fölérendeltjét az „L' kapcsolatban.

A CODASYL adatmodell fő ereje tehát az ábrázolható és az adatmodell műveleteiben felhasználható kapcsolatok gazdagsága. /A teljes műveletlista közel 100 műveletből áll [11]./ Emellett azonban rá kell mutatni arra, hogy ebben az adatbázis-filozófiában már megtalálható egy, a későbbi modellekben még nagyobb jelentőséggel bíró gondolat. Az információs modell a valós világról szerzett képünket nemcsak objektumok és kapcsolatok formájában, hanem ezek közötti logikai feltételekként is tartalmazza. Az eddigi adatmodellek ilyen feltételt vagy csak a modell jellegénél fogva tartalmaztak /a hierarchikus modellben pl. minden alárendelthez szükségszerűen tartozik egy és csak egy fölérendelt/, vagy ellenőrzéséhez programot kellett írni. A CODASYL-ban már lehetőség van az információs modellben szereplő egyes feltételek közvetlen megfogalmazására az adatmodell szintjén, olyan jellegű állítások megfogalmazásával, mint pl.:

- minden szerződéshez kötelező jelleggel tartozik megrendelő;
- csak olyan megrendelő törölhető, melynek nincs szerződése.

Ezek az állítások algoritmuskészítést takarítanak meg, pusztán azzal, hogy az adatmodellbe beépíthetők. Általános tendencia a számítástechnikában, hogy az algoritmusokat a használatos modell leíró lehetőségeinek bővítésével váltsuk ki. Esetünkben az történik, hogy az információs modell realizációja az adatmodell segítségével egyszerűbbé válik azon az áron, hogy az adatmodell fordítása a számítógép nyelvére bonyolultabb lesz. A fenti folyamat emlékeztet arra, hogy az ember saját munkájának hatékonyabbá tételére egyre bonyolultabb, de könnyebben használható és nagyobb teljesítményű gépeket hoz létre.

Mind a hierarchikus, mind a CODASYL adatbázis-filozófia megalkotásánál - noha mindkét esetben "logikai" adatmodellt, vagyis technológiától független, az információs modell által inspirált adatmodellt kívántak készíteni - a közvetlen cél hatékonyan működő számítógépes rendszer megvalósítása volt. Ez mindkét adatbázis-filozófia esetében a modelltől magától

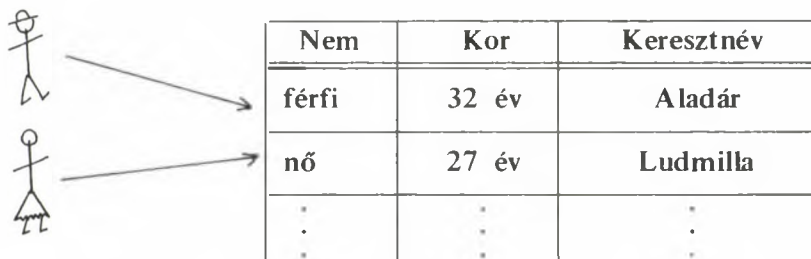
idegen, a számítógépes megvalósítás korlátait tükröző fogalmak bevezetésére vezetett. /A fizikai tárolás módját is a felhasználó adta meg, és az egyes adatokra alkalmazható műveletek nem függetlenek a tárolástól./ Ugyanakkor világos az a törekvés, hogy az adattárolást az adatmodelltől elválasszák.

Ezt az ellentmondást ismerték fel a *relációs adatbázis-filozófia* megalkotói. Immár tudatosan eltekintve a konkrét realizáció korlátaitól, egy "absztrakt" adatmodell létrehozására törekedtek. /Éppen a modell technikától független volta és fejlettsége miatt hatékony megvalósítására évekig remény sem látszott./

A relációs adatmodellek kiindulópontja, hogy a valós világ objektumai tulajdonságaik halmazával jellemezhetők. Az objektumok közötti kapcsolatokról semmiféle előzetes feltevéssel nem él, elfogadja azt a tényt, hogy az egyes objektumok a legkülönbélebb viszonyban állhatnak egymással, és általános lehetőséget ad a kapcsolatok ábrázolására.

A modell alapfogalma a *reláció*, mégpedig a matematikában használatos értelemben. A valós világ egy objektumát tulajdonságai véges és rendezett sorozataival közelítjük: az információs modell objektumait tehát a tulajdonságokból képzett relációkkal ábrázolhatjuk.

A reláció számunkra nem más, mint egy olyan táblázat, amelynek minden sora az információs modell egy objektumának a képe. A táblázat minden oszlopa egy tulajdonságot jelöl, adott oszlop adott sorában a sor által reprezentált objektumnak az oszlop által jelölt tulajdonságát találjuk meg:



A relációs modell az objektumok közötti kapcsolatok ábrázolására is relációkat használ. Ezáltal a modell teljesen homogén lesz.

Az adatmodellnek csak az objektumait tekintve úgy tűnik, hogy a legelső - rekordon és fájlban alapuló - adatmodellnél primitivebbet kaptunk: a reláció felel meg a fájlban, egy sor egy rekord, és a relációknak ráadásul rendezettségük sincs. A minőségi különbség a definiált /megvalósítható/ műveletek jellegében van.

A műveletek definíciójában nem használunk ki semmit a fizikai tárolás jellegzetességeiből, a hardver képességeiből. /Nincs kulcs, kulcs szerinti elérés, pointer, index stb./ A keresési műveletek olyan jellegűek, mint:

- keresd ki azokat a dolgozókat, akiknek a fizetése 5000 és 6000 Ft között van, életkora 35 év

feltéve, hogy a "dolgozó", mint reláció, a "fizetés" és "életkor" pedig mint e reláció oszlopai szerepelnek.

Kezdeményezhető olyan válogatás is, amelyben a szempontok egyszerre több relációt érintenek. Ilyen kérdések formájában kerülnek elő az objektumok közötti kapcsolatok.

Megjegyezzük, hogy a relációs modellben még értelmes módon feltehető kérdések - algebrai eszközökkel - háromféle alapműveletből álló sorozatokra vezethetők vissza. Ezekkel a műveletekkel és a konverziós folyamattal a felhasználó nem találkozhat. Az ő rendelkezésére álló műveletek a természetes nyelvet jól közelítő kijelentő mondatok. A nyelvben szereplő fogalmak már alig megkülönböztethetők az információs modell fogalmaitól.

A relációs adatbázis-filozófia a legutóbbi évtized kutatásainak eredménye. Sok forradalmian új eszméje közül mi a legfontosabbat emeltük ki: nem a számítógépes technológia szempontjából vizsgálja a valóságot, hanem ettől független, általános modellt ad.

ÖSSZEFOGLALÁS

Dolgozatunkban végigkövettük az információs rendszerek fejlődésének utját az emberi munkamegosztás megkövetelte legkezdetlegesebb rendszerektől egészen a jelenkor világot behálózó adatbázisáig. Nem térünk ki az információs rendszerek működésének technikai részleteire, fő célunk az adatmodell és az információs modellel reprezentált valós világ viszonyának és kölcsönhatásának vizsgálata volt.

Az általános tendencia szembeötlő: a különféle adatbázis-filozófiák mindig kompromisszumot jelentenek. Ez a kompromisszum a valós világnak az információs modellben megfogalmazódó bonyolult és komplex igényei, illetve az adott kor technológiai színvonalában realizálódó korlátok között jön létre. A technológia nemcsak az adatmodell megvalósíthatóságát korlátozza: a termelési folyamatok összetettsége, a társadalom strukturáltsága lényegesen befolyásolja magának az információs modellnek a komplexitását is. Az adatmodellek fejlődését meghatározó módon korlátozó tényező mindig a hatékony realizáció /a termelésben, illetve a társadalomban való felhasználhatóság/. Ugyanakkor az adatmodellek fejlesztése természetesen a tudomány és technika fejlődésére is hat /formális nyelvek elmélete, diszkek megjele- nése, adatbázis gépek stb./.

Az információs modell - tehát a valóságról alkotott képünk - és az adatmodell - azaz ennek tárgyiasult formája - között a fejlődés kezdeti szakaszában igen nagy eltérés volt. A technológia kezdetlegessége folytán az adatmodell fogalomrendszere nagyon primitív, ésszerűen definiálható műveletei pedig még az egyszerűbb információs modellek realizációjához is túl kezdetlegesek voltak.

Az adatmodell a fejlődés során nemcsak egyre összetettebbé - ezzel egyre hatékonyabbá - vált, hanem fogalomrendszere egyre

közelebb került az információs modellnek közvetlenül a valós világból származó fogalmához. E folyamat eredményeként az információs modellt az adathordozókon végzett közvetlen műveletekre leképező transzformációsorozat egyre nagyobb hányada vált automatizálhatóvá. Ezáltal az ember tudatos tevékenységének egyre nagyobb hányadát fordithatta tökéletesebb információs modell megalkotására, a valós folyamatok teljesebb feltárására, tökéletesebb megismerésére.

Ugyanakkor az adatmodellek fejlődése egyre tökéletesebb információs modellek létrehozására ad lehetőséget, ezzel újabb igényeket teremtve.

A várható fejlődési tendenciákat vizsgálva megállapíthatjuk, hogy a relációs modell nem lehet a fejlődés végállomása. Kialakulásával létrejött ugyan egy eszköz az információs modellek adatmodellekre való tükrözésének általános megoldására. Véleményünk szerint a relációs modell kialakulásáig vezető, egyre általánosabb fogalmakat elvonatkozható tendenciát most már egy, a relációs modell nagyon általános fogalomrendszerét specializáló tendencia követi, azaz az információs modellből szintetizáló folyamat valószínűleg egy analizáló jellegű folyamat követi. A relációs adatbázis-filozófia fogalomrendszerét osztályozva, speciális tulajdonságokkal ellátva használó modellek máris kialakulóban vannak [12], [13].

IRODALOM

- [1] Szentgyörgyi Zsuzsa: A számítástechnika műszaki fejlődése és társadalmi hatásai. *MTA SZTAKI Tanulmányok* 120, Bp. MTA SZTAKI. 1981. pp. 1-229.
- [2] Sage, D.M.: Information System: A Brief Look into History. *Datamation*, November /1968/, pp. 63-69.
- [3] Biggs, C.L., Birks, E.G., Atkins, W.: *Managing the System Development Process*. Prentice-Hall /1980/.
- [4] Ross, D.T., Schoman, K.E.: Structured Analysis for Requirements Definition. *IEEE Transactions on Software Engineering*, SE-3 /1977/. pp.6-15.
- [5] Teichroew, D., Winters, E.: Recent Developments in System Analysis and Design. *Atlante Economic Review*, November-December 1976, pp. 39-46.
- [6] *Filozófiai kislexikon*. Kossuth Könyvkiadó, Budapest, 1972.
- [7] *Műszaki lexikon*. Akadémiai Könyvkiadó, Budapest, 1974.
- [8] Neumann János: *A számítógép és az agy*. Gondolat Könyvkiadó, Budapest, 1972.
- [9] Wasserman, A.J.: Information System Design Methodology. *Journal of the American Society for Information Science*. January /1980/, pp. 5-24.
- [10] Boehm, B.W.: Software Engineering. *IEEE Transactions on Computers* C-25 /1976/ 12, pp. 1226-1241.
- [11] Olle, T.W.: *The Codasyl Approach to Data Base Management*. John Willey and Sons, New York, 1978.
- [12] Hammer, M., McLeod, D.: Database Description with SDM: A Semantic Database Model. *ACM TODS* 6 /1981/ 3, pp. 351-386.
- [13] Markowitz, H.M., Malhotra, A., Pazel, D.P.: The EAS-E Application Development System: Principles and Language Summary. *CACM* 27 /1984/ 8, pp. 785-799.

Data models and the data base philosophy

This paper intends to characterize the way in which data models reflect the real world. Some concepts /data model, information model, e.t.c./ are introduced for this purpose, and a model describing the information system development process is presented in this terminology. Characteristic features of some data models - from pre-computer ones to the relational - are investigated analyzing their relationship with the sociological-technological background.

1985-BEN EDDIG MEGJELENTEK:

- 166/1985 Radó Péter: Információs rendszerek számítógépes tervezése
- 167/1985 Studies in Applied Stochastic Programming I.
Szerkesztette: Prékopa András
- 168/1985 Böszörményi László - Kovács László - Martos Balázs
Szabó Miklós: LILIPUTH
- 169/1985 Horváth Mátyás: Alkatrészgyártási folyamatok automatizált tervezése
- 170/1985 Márkus Gábor: Algoritmus mátrix alapú logaritmus kiszámítására kriptográfiai alkalmazásokkal
- 171/1985 Tamás Várady: Integration of free-form surfaces into a volumetric modeller
- 172/1985 Reviczky János: A számítógépes grafika terület- kitöltő algoritmusai
- 173/1985 Kacsukné Bruckner Livia: Mozgáspálya generálás bonyolult geometriájú felületek 2 1/2D-s NC megmunkálásához
- 174/1985 Bolla Marianna: Mátrixok spektrálfelbontásának és szinguláris felbontásának módszerei.
- 175/1985 Hannák László - Radó Péter: Adatmodellek, adatmérés-filozófiák.



1985 DEC 27

1111