

# Értékünk AZ **EMBER**

Humán erőforrás-fejlesztési Operatív Program



Dr. Sziray József  
Gaul Géza  
Dr. Égertné dr. Molnár Éva

## VEZETŐI INFORMÁCIÓS RENDSZEREK

Készült a HEFOP 3.3.1-P.-2004-09-0102/1.0 pályázat támogatásával.

Szerzők: Dr. Sziray József  
egyetemi docens  
Gaul Géza  
egyetemi adjunktus  
Dr. Égertné dr. Molnár Éva  
főiskolai docens

Lektor: Dr. Kovács János  
egyetemi docens

# A dokumentum használata

## Mozgás a dokumentumban

A dokumentumban való mozgáshoz a Windows és az Adobe Reader megszokott elemeit és módszereit használhatjuk.

Minden lap tetején és alján egy navigációs sor található, itt a megfelelő hivatkozásra kattintva ugorhatunk a használati útmutatóra, a tartalomjegyzékre, valamint a tárgymutatóra. A ◀ és a ▶ nyilakkal az előző és a következő oldalra léphetünk át, míg a Vissza mező az utoljára megnézett oldalra visz vissza bennünket.

## Pozicionálás a könyvjelzőablak segítségével

A bal oldali könyvjelző ablakban tartalomjegyzékfa található, amelynek bejegyzéseire kattintva az adott fejezet/alfejezet első oldalára jutunk. Az aktuális pozíciókat a tartalomjegyzékfában kiemelt bejegyzés mutatja.

## A tartalomjegyzék használata

### Ugrás megadott helyre a tartalomjegyzék segítségével

Kattintsunk a tartalomjegyzék megfelelő pontjára, ezzel az adott fejezet első oldalára jutunk.

### Keresés a szövegben

A dokumentumban való kereséshez használjuk megszokott módon a Szerkesztés menü Keresés parancsát. Az Adobe Reader az adott pozíciótól kezdve keres a szövegben.

# Tartalomjegyzék

<b>Bevezetés</b> .....	<b>6</b>
<b>1. A vezetői információs rendszerek szerepe</b> .....	<b>9</b>
1.1. Az alkalmazás módja és iránya.....	9
1.2. Felhasználási célok.....	11
1.3. A megfigyelési paraméterek meghatározása.....	11
1.4. Feldolgozási módok a VIR-ekben.....	12
1.5. Elérendő célok .....	15
1.6. Egy VIR létrehozása.....	16
1.7. Egy VIR életciklusa .....	17
1.8. Alkalmazási esetek .....	17
1.9. Tanácsadás: Andersen Consulting (UK) .....	17
1.10. Kereskedelem: Pepsi Co. – Frito-Lay (USA).....	18
1.11. Felhasználás üzleti tervekhez .....	19
<b>2. A VIR-ek létrehozási folyamatai</b> .....	<b>21</b>
2.1. A VIR-ek létrehozásának szempontjai.....	21
2.2. Veszélyek feltárása.....	22
2.3. Az üzletmenet ismerete.....	23
2.4. Fokozatos bővítés, fejlesztés.....	23
2.5. A csoportmunka szervezése.....	24
<b>3. Felhasználási folyamatok</b> .....	<b>25</b>
3.1. Adatbázis-struktúrák .....	25
3.2. Egy hazai vegyi üzem eladási modellje .....	28
3.3. Többdimenziós (multidimenzionális) adatstruktúrák.....	33
3.4. Adatkeresési folyamatok a többdimenziós struktúrában .....	33
3.5. Az OLAP végrehajtása.....	35
3.6. Szakértői rendszerek használata VIR-ekben.....	38
<b>4. Stratégiai menedzsment és informatikai támogatás</b> .....	<b>43</b>
4.1. Stratégiai folyamatok .....	43
4.2. A stratégiai menedzsmentben használt módszerek .....	47
4.3. A menedzser-barát informatika .....	48
4.4. Algoritmusok adat-elérési, szűrés, elemzési célra.....	51
4.5. Mutatók, mutatószámok .....	52
4.6. A képi megjelenítés (vizualizálás) módszere .....	57

<b>5. Az OLAP technológia.....</b>	<b>74</b>
5.1. Az OLAP Codd-féle kritériumai .....	74
5.2. Az OLTP és OLAP rendszerek összehasonlítása.....	76
5.3. A többdimenziós (multidimenzionális) adatmodell.....	78
<b>6. Adatbányászás .....</b>	<b>94</b>
6.1. A tudásfeltárás folyamata és az adatbányászás .....	94
6.2. Az adatbányászás alkalmazási lehetőségei.....	96
6.3. Adatbányászati módszertanok és lekérdező nyelvek.....	100
6.4. Társítási szabályok meghatározása .....	107
6.5. Csoportosítási (kluszterelemzési) algoritmusok .....	111
6.6. Osztályozási algoritmusok.....	115
6.7. Adatbányászati szoftverek.....	119
<b>7. A Comshare rendszer .....</b>	<b>120</b>
7.1. A termék első generációjának kialakítása.....	120
7.2. A Commander funkciói .....	121
7.3. Néhány jellemző Comshare VIR-alkalmazás.....	132
7.4. A Comshare második generációs termékei .....	133
<b>Irodalomjegyzék.....</b>	<b>136</b>
<b>Rövidítések.....</b>	<b>138</b>

## Bevezetés

Az intézmények, hivatalok hatékony működtetésében, a cégek, vállalatok világában, ill. a közöttük folyó piaci versenyben egyre fontosabbá válik az általuk felhasznált informatikai rendszerek minél jobb hatásfokú kihasználása. Ez igen erősen jelentkezik a vállalatok munkaerő-keresletében, ami az alábbi szakterületekre irányul:

- rendszerfejlesztők,
- rendszergazdák,
- kontrolling-osok,
- vezetői informatikai tanácsadás-konzultálás,
- informatikai rendszerek auditálása,
- informatikai projektek minőségbiztosítása,
- üzleti-vállalati folyamatok újjászervezése stb.

Ennek az a következménye, hogy folyamatosan növekszik a vállalati informatikában járatos, széles látókörű, rendszerekben gondolkozni tudó informatikai szakemberek iránti igény. Ez az expanzív tendencia egyelőre beláthatatlan ideig folytatódni látszik. A felsőoktatásban folyó informatikusképzésben új, korszerű tananyagokkal kell ezt az igényt kielégíteni.

A Széchenyi István Egyetemen a vállalati informatika területén már régóta folyik oktatás. Ehhez kapcsolódóan jelentek meg itt a „Vállalati információs rendszerek I.” és a „Vállalati információs rendszerek II.” című egyetemi jegyzetek, amelyek az egyes intézmények által felhasznált információs rendszerek üzemeltetésével, használatával, szervezésével kapcsolatos ismereteket foglalják magukban.

A szóba jöhető intézmények lehetséges köre:

- közigazgatási,
- hatósági,
- oktatási,
- termelési,
- kutatási-fejlesztési,
- kereskedelmi,
- szolgáltatási stb.

Felhasználási módok:

- üzleti műveletek támogatása,
- vezetői döntések támogatása,
- stratégiai támogatás a piaci versenyben.

Ebben a tankönyvben a vállalati informatika egy fontos, külön kezelhető területével, a vezetői döntéseket támogató informatikai rendszerekkel, az ún. Vezetői Információs Rendszerekkel (VIR-ek) foglalkozunk. Az olvasó itt megismerkedhet az ilyen típusú szoftver eszközök felhasználási módjával, alkalmazási irányjaival, a VIR-ek funkciójával, szerepükkel a vállalat üzleti folyamatainak irányításában, továbbá magukkal a vezetők által végzendő felhasználási folyamatokkal.

A tankönyv 1. fejezete ismerteti a VIR-ek, valamint a hozzá kapcsolódó fontosabb fogalmaknak a definícióját. Ugyanitt kerül sor a VIR-ek általános tulajdonságainak, valamint felhasználási lehetőségeinek az áttekintésére.

A 2. fejezetben egy VIR-nek a vállalaton belüli kiépítésével, létrehozásával foglalkozunk. Ehhez kapcsolódóan ismertetjük a létrehozás és üzemeltetés szempontjait és lehetséges problémáit.

A 3. fejezetben először a VIR-ek adatfelépítési típusait tekintjük át, majd a különböző adatstruktúrákhoz kapcsolódó felhasználási módokat mutatjuk be, kisebb alkalmazási esetekkel illusztrálva. Ebben a fejezetben foglalkozunk még a számítógépes üzleti intelligenciának, ezen belül a szakértői rendszereknek a döntéstámogatásban való felhasználásával.

A 4. fejezet áttekinti a vezetői tevékenységnek, és – ebből fakadóan – a vezetői információs rendszereknek a helyét a vállalati-intézményi döntéshozatali hierarchiában, és ennek a legfontosabb következményeit. Mivel erre a tevékenységterületre a feladatoknak és információknak az alsóbb szinteken lévőkhöz képest lényegesen kisebb fokú strukturáltsága jellemző, és az ezen a szinten hozott döntésekben lényegesen nagyobb a döntéshozó intuitív hozzájárulása, ezért itt másként, és sokkal nagyobb mértékben játszanak szerepet az ember információ-kezelési és feldolgozási adottságai, mint a vállalati információs rendszerekben általában. Ezt, ennek következményeit tárgyaljuk ebben a fejezetben.

Az 5. fejezet a relációs adatbázisokon, az azok adataiból felépített adatáruházakon (adattárházakon, Data Warehouse) végezhető elemzések céljára kidolgozott on-line elemző feldolgozásról (On-line Analytical Processing, OLAP) szól. Az OLAP legfontosabb ismérve, hogy multidimenzionális

adatstruktúrát használva lehetővé teszi az adatok gyors és rugalmas lekérdezését, majd ezt követően az adatok analízisét.

A 6. fejezet az adatbányászásról (Data Mining) szól. Az adatbányászás olyan adatelemzési folyamat, amely nagy adatbázisokból rejtett, előzőleg nem ismert, hasznos információkat tár fel. A rejtett összefüggések felismerése és kinyerése mellett igen fontos, hogy adatbányászási módszerekkel szakértői döntések automatizálhatók, ami lehetővé teszi emberi erőforrás megtakarítását és a kérdésekre adott válaszok reakcióidejének lerövidítését.

A 7. fejezetben az amerikai Comshare cég által kifejlesztett és világszerte elterjedt vezetői információs rendszerek két generációjának felépítését és működést ismertetjük, részletesebben bemutatva a szoftverek funkcióit és üzemeltetési lehetőségeit.

A könyv végén a megíráshoz felhasznált irodalom felsorolása található.

Ez a tankönyv elsősorban azoknak az egyetemi és főiskolai hallgatóknak ajánlható, akik olyan informatikai képzésben részesülnek, amibe beletartozik a vállalati információs rendszerek működésének és használatának megismerése.

Budapest–Győr, 2007. szeptember 15.

A szerzők



# 1. A vezetői információs rendszerek szerepe

## 1.1. Az alkalmazás módja és iránya

Az éles piaci verseny és a gazdasági döntéshozók megnövekedett felelősége fokozottan előtérbe hozta Magyarországon is a naprakész információkra épülő, gyors és megalapozott döntések jelentőségét. Az ún. *Vezetői Információs Rendszerek (VIR-ek)* elsősorban felső szintű döntéshozók, pénzügyi szakemberek és üzletemberek számára nyújtanak hatékony támogatást döntéseik előkészítése során.

A siker záloga, hogy a döntéshozók szempontjait és gondolkodásmódját figyelembe véve hozták létre a VIR-eket, amelyek használatában egyébként nincs szükség számítástechnikai végzettségre.

A VIR (angol elnevezéssel: *Executive Information System: EIS*) meghatározása a következő:

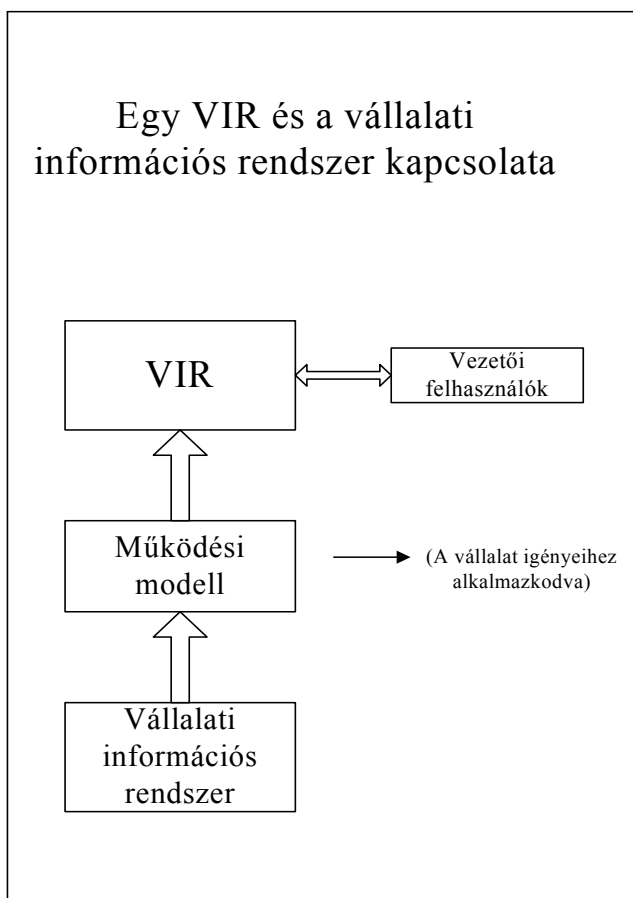
*Vezetői információs rendszer (VIR):* Olyan szoftver, amely egy vállalat információs rendszerére épül, s ebből egy előre kidolgozott működési modell alapján szűri ki a vezetők számára azt az információt, amire az üzleti folyamatok irányításához leginkább szükségük van. A VIR soha nem dolgozik bele a vállalati adatbázisba, onnan csak kiválasztja az adatokat.

A VIR a vállalati rendszer „tetejére” van helyezve, és a vezetési döntésekhez szükséges információt szűri ki, szelektálja az adatbázisból, és ezt „tálalja” a menedzsment számára. Ez az információ könnyen áttekinthető, jól kezelhető formában áll rendelkezésre.

Az itt leírtak áttekinthető szemléltetése az 1.1. ábrán látható.

A VIR-ek fejlesztése hosszú tapasztalat és széles körű számítástechnikai szaktudás, hatalmas fejlesztési befektetés eredménye. Az ilyen rendszereket egy-egy SW-cég hozza létre és viszi ki a piacra, forgalmazás céljából. Mindehhez a legkorszerűbb menedzselési, elemzési, döntési, irányítási know-how-t használják fel, és építik be ugyancsak modern, hatékony számítástechnikai megoldásokba.

Ilyen például a *Comshare Co. (USA)* cég által 1993-ban kifejlesztett új, *Commander* nevű VIR, amelynek az első verziós fejlesztési költsége 10 millió dollár volt. (Mai áron kb. 2 milliárd forint).



**1.1. ábra:** Egy VIR és a vállalati információs rendszer kapcsolata

A VIR-ek főként arra szolgálnak, hogy az üzleti élet felső vezetői számára gyors és hatékony analízisi lehetőséget biztosítsanak, kritikus összesített adathalmazok körében. Mindehhez óriási adatfeldolgozási kapacitást, és jól áttekinthető vizuális (grafikus) környezetet nyújtanak. Az eddigi felmérések szerint a feldolgozott információ jó része pénzügyi, de ezen kívül egyre elterjedtebben kezelnek más jellegű információt is, mint például gyártási, eladási, marketing adatokat, külső híreket, piaci trendeket stb.

## 1.2. Felhasználási célok

Egy amerikai felmérés eredménye 100 megkérdezett vállalat esetében (1990.), arra vonatkozóan, hogy milyen célokra használják a VIR-t:

- 85%: Pénzügyi adatok figyelése.
- 67%: Stratégiai adatok figyelése.
- 45%: Marketing és eladási adatok figyelése.
- 40%: Személyzeti adatok (emberi erőforrás – human resources) figyelése.
- 27%: Piaci verseny adatainak figyelése.
- 20%: Gyártási adatok figyelése.

Az eltelt időszak trendjei azt mutatták, hogy lényegében ma is ugyanezek a felhasználási célok érvényesülnek a piaci világban.

## 1.3. A megfigyelési paraméterek meghatározása

Egy sikeres VIR létrehozása és bevezetése megköveteli a cég információ-szükségletének szigorú elemzését, precíz kialakítását. Ez azt jelenti, hogy a vezetőknek tudniuk kell, hogy melyek azok az adatok, amelyek a vállalat sikerét vagy kudarcát jelzik. Egy ilyen elemzés stratégiai vagy akár szervezeti változáshoz is vezethet. Ilyen változás lehet például:

- egy új termék kifejlesztése,
- egy új szolgáltatás bevezetése.

Ebben a folyamatban igen nagy jelentősége van az innovációnak, aminek az eredménye lehet:

- új eljárás,
- új termék,
- új piac.

A VIR-ek eredményes használata szempontjából igen fontos az, hogy a vezetők milyen mutatókat, paramétereket vonnak be a folyamatos megfigyelésbe. Ezek célszerűen olyan adatok kell hogy legyenek, amelyek a vezető által felügyelt területen a vállalat működési állapotát a legjobban írják le, a legjobban jellemzik.

Ilyen jellegű figyelési paraméterek, amelyek általános fogalomként is használандók, a következők:

**Kritikus sikertényező (KST) (Critical Success Factor – CSF):**

Azok a mutatók, amelyek előírt értékeinek elérése a vállalat fő célja. Ilyen mutatók lehetnek például:

- árbevétel,
- tiszta nyerség,
- költségszint,
- likviditási szint,
- piaci részesedés aránya stb.

**Kulcsfontosságú teljesítménymutató (KTM)****(Key Performance Indicator – KPI):**

Relatív mutató, ami azt jelzi, hogy milyen mértékben, arányban tudták elérni a kitűzött célokat, trendeket. Például:

- a vásárlók elégedettségi szintje, a panaszok, reklamációk, ill. dicsérő észrevételek arányában,
- piaci részesedés változási aránya egy negyedévre nézve

**Kiugró eltérések (kivételek) kezelése (Exception Handling):**

Kijelölt paraméterek automatikus figyelése, előzetesen megadott határértékek (minimum, maximum), ill. tűrési sávok (-tól, -ig) alapján.

Ha egy paraméter az előírt értéktől egy határon túl a megadott irányban eltér: piros értékkijelzés. Figyelmeztetés (Early Warning) egy tűrési határon belül: sárga értékkijelzés. Rendben levő érték (OK): zöld.

**1.4. Feldolgozási módok a VIR-ekben**

Egy VIR-ben a vezetők végeredményben a vállalat adatbázisához férnek hozzá, egy külön e célra létrehozott számítási, modellezési folyamat révén. Ilyen alapvető és jellegzetes folyamatok az alábbiak:

**Függőleges leásás az adatbázisban****(zoom-olás, drilling, adatbányászás, data mining):**

Lefelé való navigálás a hierarchikusan szervezett és egymás alatti rétegekkel rendelkező adatbázisban.

Az ilyen keresésnek az a célja, hogy egy felső szinten jelentkező problémának, irányzatnak, vagy jelenségnek meg lehessen találni a mélyebben rejlő okát, eredetét.

Például, egy országos méretű áruházi lánc regionális szinten jelentkező piacvesztésének okait meghatározni. Időbeli eladási folyamatok vizsgálata is lehetséges ilyenkor.

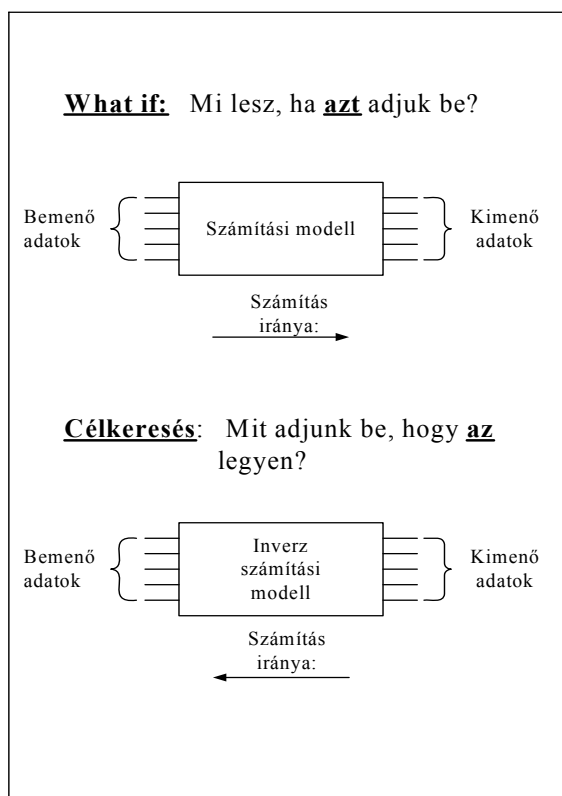
### „Mi lenne ha” típusú elemzés (What If Analysis):

Különböző célú modellezési kísérletek végezhetők el ezáltal. Feltételezett adatok előírásával a rendszerrel kiszámítatjuk, hogy milyen eredmények lennének várhatóak.

### Célkeresés (Goal Seeking):

Ez a lehetőség az előbbi folyamat inverzét jelenti: Ilyenkor az elvárt értéket (vagy értékeket) írjuk elő, és a rendszer azt számítja ki, hogy az adott értékek (célok) eléréséhez milyen értékű bemenő paraméterek szükségesek. (Inverz modellezés.)

Az itt leírt két megközelítés számítási sémáit az 1.2. ábra mutatja be.



1.2. ábra: Tervezési számítások két irányban

### Többdimenziós adatbázis-kezelés:

A dimenziók száma azt fejezi ki, hogy egy adat hány másik adattól, paramétertől függ.

Például: **Összesített eladási volumen:**

- hely,
- idő,
- termék.

A példában ez három dimenzió, amit úgy kell értelmezni, hogy az *összesített eladási volumen* a fenti három adat különböző konkrét érték kombinációjától függően vesz fel különböző értékeket. Ez a kapcsolat tulajdonképpen matematikailag értelmezhető többváltozós diszkrét függvényként is, mégpedig az esetünkben három diszkrét változóval.

Az együttesen kezelt dimenziók számától függően, ezekből akárhány dimenziós *metszet* készíthető, és eszerint vetíti ki a VIR az adatcsoportokat, rendezett formában. Matematikai értelmezésünk szerint egy metszet a diszkrét változók egy adott részhalmazának rögzített értéke mellett felvett függvényértékek halmaza, ahol a nem rögzített változók különböző értékekkel vesznek részt a metszet kialakításában.

### OLAP (On-Line Analytical Processing):

Többdimenziós adatnyomozás a vállalati információs rendszer folyamatos üzemelése alatt, ahol a feldolgozási idő gyakorlatilag nem függ a dimenziók számától. Az adatnyomozás folyamatában a vezető szándékától függően változnak a rögzített és a nem rögzített adatok, ami különböző metszetekben teszi lehetővé az adatok közötti lényeges összefüggések megállapítását.

Az OLAP- rendszereket egy későbbi, önálló fejezetben részletesen ismertetjük. Itt csak a jelen fejezet gondolatmenetéhez szükséges legfontosabb ismereteket emeljük ki.

### E. F. Codd (USA) értékelési kritériumai az OLAP rendszerekre vonatkozóan:

Az alábbi területeken kell hogy megfeleljen a Codd által megadott szabályoknak egy valódi OLAP rendszer:

1. Multidimenzionális kezelési képesség.
2. Átláthatóság.
3. Elérhetőség.
4. Konzisztens jelentéskészítési képesség.

5. Kliens/szerver architektúra.
6. Egységes dimenziómodellezés.
7. A nem teljesen kitöltött (hiányos) mátrixok dinamikus kezelése.
8. Többfelhasználós üzemelés támogatása.
9. A dimenziók közötti műveletek korlátozás nélküli végrehajthatósága.
10. Intuitív (közvetlen, segédeszköz nélküli) adatmanipuláció lehetősége.
11. Rugalmas jelentéskészítés.
12. A dimenziók és az aggregációs szintek száma korlátozás nélküli.

Megjegyzések:

- A Codd-féle kritériumok a piacon elérhető OLAP rendszerek funkcionalitásának, hatékonyságának, ill. teljesítményének összehasonlítására is alkalmazhatók.
- A felmérések szerint a felhasználóknak maximálisan 19 konkurrens adatdimenzióra lehet szükségük. (Itt arról is szó lehet, hogy a dimenziók emberi áttekinthetése ütközik felső korlátba.)
- Általános egyetértés van abban, hogy egy komoly teljesítményű OLAP eszköznek legalább 15 dimenziót kell tudnia kezelni, amely szám lehetőleg 20-ig mehessen fel.

## 1.5. Elérendő célok

Egyes szakértők szerint egy VIR szerepét úgy lehet elképzelni, mintha egy háborús vezérkari irányítószobában lennénk, és a számítógépes rendszer abban segítene, hogy a rengeteg beérkező információból kigyűjtse azokat, amelyek a gyorsan változó helyzethez való igazodás döntéseit segítik.

Felmérés: 1990 – Washington: Egy VIR-konferencia résztvevői körében:

### A VIR használatával elérendő célok fontossági sorrendje:

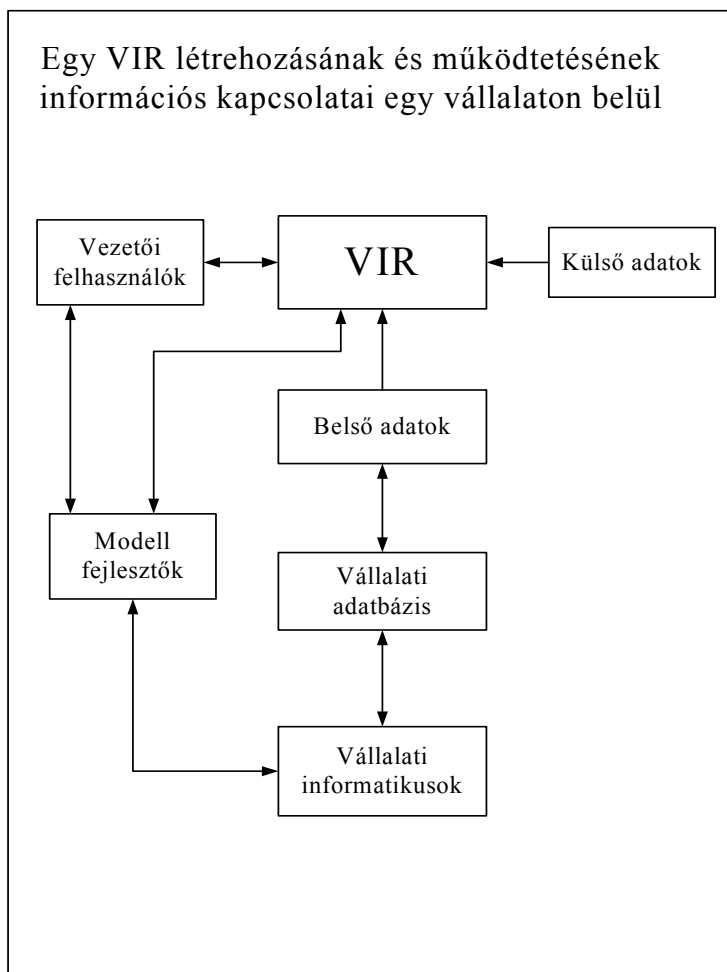
1. A menedzseri jelentések színvonalának javulása.
2. Több időszerű adat.
3. Stratégiai tervezés javulása.
4. Megbízható adatbázis kialakítása. (Ez visszacsatolás.)
5. Pénzügyi irányítás javulása.
6. A vállalati szervezet jobb megértése.
7. A gazdasági és piaci adatok minőségének javulása.
8. A termékminőség és a szolgáltatás színvonalának javulása.
9. A piaci versenyanalízis minőségének javulása.
10. Vezetői team-ek közötti kommunikáció javulása.

## 1.6. Egy VIR létrehozása

A VIR-t képviselő szoftver-terméket a megvásárló vállalatnál az ottani céloknak és felhasználási igényeknek megfelelően a vállalat döntési folyamataihoz kell igazítani. Ez két lépésben történik:

1. A működési modell megtervezése, kialakítása.
2. A modell beintegrálása a VIR és a vállalati információs rendszer közé.

A létrehozás és működtetés folyamata az 1.3. ábrán követhető végig.



1.3. ábra: Egy VIR létrehozásának és működtetésének folyamata



## 1.7. Egy VIR életciklusa

Mint minden egyes szoftver terméknek, egy VIR-nek is természetes *életciklusa* van. Ez általában 5–8 évre tehető, és a következő fázisokból áll:

- A prototípus és a kezdeti kiépítés előállítása: 6–8 hónap,
- Alkalmazás és továbbfejlesztés: 18–24 hónap,
- Felhasználók számának növelése és az alkalmazás megszilárdítása: 18–24 hónap,
- Komolyabb átalakítás az újabban felmerülő igények alapján: 18 hónap,
- Újraüzemeltetés: 12 hónap.

## 1.8. Alkalmazási esetek

Eddigi főbb alkalmazási területek nyugaton:

- Vegyipar.
- Tanácsadás.
- Pénzügyi szolgáltatás.
- Gyártás.
- Könyvkiadás.
- Kereskedelem.
- Szállítás.

A következőkben két esettanulmányt tekintünk át. Ezekben jellegzetes VIR-alkalmazásokat írunk le vázlatosan, kiemelve, hogy azok miért váltak fontossá és kifizetődővé az illető cégek számára.

## 1.9. Tanácsadás: Andersen Consulting (UK)

(Esettanulmány)

*A cég fő tevékenység:* Pénzügyi folyamatok auditálása, vezetői tanácsadás.

*VIR:* Comshare – COMMANDER.

18.000 alkalmazottja van a világon.

Tevékenységi területek a VIR-ben:

- Pénzügyek,
- költségvetés,
- üzleti terv,
- munkák figyelése,
- személyzeti ügyek.

Itt a hangsúly a konzultációs munkák figyelésén volt. Ezen belül trendeket figyeltek, a partnerek, társak, menedzserek és konzultánsok teljesítményét kísérték figyelemmel.

Itt az elszámolás alapja, vagyis a kulcsfontosságú teljesítménymutató:

### **KTM: a tanácsadási idő.**

(Például, egy átlagos értékszám: 800 font/nap/fő.)

Előrejelzéseket is tudtak készíteni a rendszer felhasználásával, egy-két hónapra.

Míndez az emberi erőforrások optimális elosztását, szervezését, ebből eredően az optimális kihasználását tette lehetővé.

(Partner: Ügyfélként kezelt vállalat, cég, intézmény, ahol a tanácsadási tevékenységet folytatják.)

Az Andersen Consulting bevétele:

**évente 1,8 millió font/partner.**

Egy összehasonlító adat egy másik hasonló profilú céggel:

Price & Waterhouse: 0,98 millió font/partner.

Míndezt a VIR hatékony felhasználása segítette elő.

## **1.10. Kereskedelem: Pepsi Co. – Frito-Lay (USA)**

(Esettanulmány)

*A cég működési területe: Gyorsételek forgalmazása.*

Tevékenységi területek a VIR-ben:

- Menedzsment,
- pénzügyek,
- költségvetés,
- üzleti terv,
- eladási adatok gyűjtése,
- vevőszolgálat,
- munkák figyelése,
- személyzet nyilvántartása,
- versenytársak figyelése,
- marketing.

Felhasználói helyek száma az USA-ban: 600.

VIR: Comshare – COMMANDER.

Ez a cég rendelkezik a legjobban dokumentált VIR-rel a világon.

Alkalmazás: 1986 óta. (Ez azt jelenti, hogy hosszú gyakorlatot szereztek különböző döntéstámogató számítógépes rendszerek felhasználása területén.)

### **Eladási információ gyűjtése:**

Az USA-ban 10.000 szállító-eladó mobil számítógépbe gyűjti az adatokat, majd a munkanap elteltével modemen, vagy hálózaton keresztül továbbítja a központba, Dallasba. (Ezzel az adminisztrációs munka csökkent heti 2 órával, személyenként.)

Ennek az információnak a megfelelő felhasználása lehetővé tette a körzetek áruellátásának optimális szervezését is.

A cég ezen kívül még rendszeresen beszerzi egy önálló információszolgáltató vállalattól a saját és a versenytársakra vonatkozó vásárlási adatokat. (Saját: heti gyakorisággal. Versenytársak: havi gyakorisággal.)

A megrendelésre begyűjtött adatokat beviszik a VIR-be és rendszeresen elemzik.

Ez a rendszer lehetővé tette a cég átstrukturálását, és szerintük nélküle nem tudnák tartani vezető pozíciójukat a piacon.

## **1.11. Felhasználás üzleti tervekhez**

Az *üzleti terv* egy vállalat, cég előre elkészített terve egy adott időszakra, a várható árbevételekre, üzleti eredményekre, ill. kiadásokra, valamint költségekre vonatkozóan. A költségekbe a beruházások, a bérek és az adóterhek is beletartoznak, az esetleges hitelek kamatterhével együtt. Az üzleti terv a kijelölt időszak alatt időbeli szakaszokra, például hónapokra, vagy negyedévekre van bontva. A terv időszakja általában egy vagy két-három év szokott lenni, attól függően, hogy a menedzsment milyen mértékű távlatokban kíván tervezni. Természetesen, lehet értelme hosszabb etapban is előrevetíteni az elképzeléseket.

### **1. Felhasználás a tervek készítésében:**

- Adatbázishoz való kapcsolódás.
- Modellezés.
- „Mi lenne ha” típusú vizsgálatok. Lehetséges alternatívák készítése és összehasonlítása.
- Célkeresési vizsgálatok (például a profit maximalizálása).

## 2. Felhasználás a terv megvalósulásának követése során:

- Leásásos elemzések (többdimenziós keresések).
- Kiugró eltérések okainak keresése.
- Tervadatok és tényadatok összehasonlítása.
- Trendelemzések.

*Továbbá:* Az 1. alatti lehetőségek itt is érvényesek maradnak. Ugyanis a megvalósítás során is szükségessé válhat menet közben résztervek készítése, ill. pénzügyi modellezés végzése.

## 2. A VIR-ek létrehozási folyamatai

### 2.1. A VIR-ek létrehozásának szempontjai

Egy VIR létrehozása egy vállalaton belül általában komoly beruházási költséget jelent (több százezer dollárt is kitehet hardverrel együtt). Ebbe beletartozik a vezetői modell kialakítása, és az ehhez tartozó testreszabási fejlesztések elvégzése, valamint a szükséges rendszerszervezés, oktatás és beüzemelés megvalósítása. Egy ilyen méretű beruházás nyilván csak akkor éri meg a vállalatnak, ha az alkalmazásával olyan költségmegtakarítást, ill. bevételnövekedést érnek el, ami kellően megnövekedett hasznot eredményez. A nyugati világban az éles piaci verseny készíti a vállalatokat ilyen szoftverek alkalmazására. Ez a folyamat, még ha lassan, de mindenképpen érvényesül Magyarországon is, és egyre inkább erősödni fog.

Mint tudjuk, itt rendkívül összetett rendszerekről van szó, amelyek hatékony kihasználása nagy mértékben függ a vállalati környezettől, technológiától, üzletviteltől, valamint a felhasználás módjától. Általában elmondható, hogy a VIR-ek sikeres alkalmazása leginkább a következő tényezőktől függ:

- specifikálás,
- tervezés,
- szervezeti fogadtatás,
- menedzsment,
- a VIR fejlesztési folyamata,
- technológia.

A legfontosabb feladat a vezetői információs szükséglet helyes definiálása. A sikertelen projektek felelősei gyakran azzal mentegetőznek, hogy: „a mi vezetőink nem tudják, hogy milyen információra van szükségük.” Ez azonban többnyire nem igaz: a vezetők igenis tudják, mire van szükségük. Olyan információra, ami a vállalat céljainak az elérésére használandó fel, és a mindennapi vezetői tevékenységhez bizonyul hasznosnak.

Jóllehet egy vezetői információs rendszer a vállalat speciális igényeihez kell hogy igazodjon, mégis felsorolhatunk olyan, a vezetők számára fontos követelményeket, amelyek általánosnak tekinthetők:

- Azonnali megbízható információ, amely úgy van feldolgozva, hogy a leghűbb általános képet nyújtsa a vállalatról, bármelyik időpontban, és amely kijelöli a lehetőségeket és a veszélyeket is.
- Összetett pénzügyi és működési információ, amely úgy van rendszerezve, hogy megkönnyítse az összehasonlítást előző időszakokkal, előző költségvetésekkel, és amely lehetővé teszi, hogy trendeket, ill. kivételes eseteket lehessen találni.
- Gyorsan kapott összesített adatok előrejelzésekhez, valamint a változó üzleti, piaci helyzetre való reagáláshoz.
- Nem pénzügyi jellegű, de a vállalat jövőbeni helyzetét befolyásoló információk. Ezek az alábbiak:
  - ügyfelek meglegedettsége,
  - termékminőség,
  - szolgáltatási minőség,
  - személyzeti állomány,
  - versenytársak helyzete,
  - piaci aktivitás.
- A vállalaton belül széles körben érdemes megosztani az információt, ezzel a vezetők irányítási köre növekszik, javul a betekintésük a beosztottak munkájába, másrészt pedig több ember lesz részese a siker érzésének.

## 2.2. Veszélyek feltárása

A tipikusan jelentkező problémák körébe tartoznak a következők:

- A szervezeti változásokat nehezen követi a VIR.
- Nem megfelelő projektvezetés.
- Gyenge kommunikáció (a felhasználók és az információ-adók között).
- Sokszor maguk a vezetők is elzárkóznak az információ-adás elől, azért, hogy „védjék birodalmukat”.
- Sokan csodát várnak a vezetői információs rendszertől, és ha az nem felel meg a várakozásnak, csalódottan elfordulnak tőle.
- Nagyon fontos, hogy a legbefolyásosabb vezetők támogassák a projektet. Ennek hiányában minden tönkre mehet.
- Ugyancsak fontos, hogy a felhasználók csoportja érdekeltté legyen téve a vállalaton belül.

- Három elem kapcsolata lényeges egy VIR-nél:
  - a felhasználó vezető,
  - az információ,
  - a technológia.
- Egy VIR sokszor azért fullad kudarcba, mert azt hiszik róla, hogy nem szolgáltat igazán értékes információkat, nem segít az üzleti problémák megoldásában. Sokan jobban hisznek a papíron tált adatoknak, mint a képernyőn levőnek.

### 2.3. Az üzletmenet ismerete

A rendszer sikeréhez az kell, hogy valami újat produkáljon: például trend-diagramok ügyes kezelését adja. Nem szabad sok fölösleges pénzügyi információt belevinni, pusztán csak azért, mert az rendelkezésre áll. Az információnak a kulcsfontosságú teljesítménymutatókat (KTM) kell adni, amelyek az üzlet kritikus sikertényezőivel (KST) vannak összefüggésben.

Végül is, azt lehet mondani, hogy mindazok az adatok, amelyek az *üzleti sikert* befolyásolják, benne kell hogy legyenek a rendszerben.

Röviden: érteni kell az üzletmenethez.

Idetartozó fontos információk:

- üzleti terv,
- éves jelentés,
- a vállalatra vonatkozó sajtóanyagok,
- háttér-információ a vállaltról és az iparáról,
- vállalati szervezet,
- vállalati termékek,
- vállalati szolgáltatások.

### 2.4. Fokozatos bővítés, fejlesztés

Mindegyik VIR egy viszonylag egyszerű alkalmazási feladatcsoporttal indul. Az egyes alkalmazásoknak is megvan a maguk életciklusa. Ha egy alkalmazás ennek során eléri futási maximumát, nem szabad megvárni, míg a leszálló ágba kerül, hanem új, jelentősen javított, bővített alkalmazást kell rá építeni, vagy esetleg teljesen újat indítani helyette.

Az egyes alkalmazások sorozatos javítása, bővítése, továbbfejlesztése végül is egy komplex, kifinomult rendszer kiépüléséhez vezet, amely stratégiai célokra használható, és jelentős üzleti sikert eredményez.

Vannak olyan vállalatok, amelyek már a harmadik, negyedik VIR-építési folyamatuknál tartanak. Azon dolgoznak, hogy minél jobban használják fel a technológiájukat és információjukat. Egy-egy új VIR-építés alapulhat ugyanazon a szoftver rendszeren, vagy pedig egy újabb beszerzésű terméken.

## 2.5. A csoportmunka szervezése

A VIR-ek sikeres alkalmazása lényegesen függ annak a csoporttevékenységnek, csoportmunkának a hatékonyságától, amit a rendszer használói fejtenek ki.

A számítástechnikai eszközök felhasználásával végzett csoporttevékenységet *groupwork-nek*, a munka eredményét pedig *groupware-nek* nevezzük.

*A groupware eredményességének szempontjai:*

1. A felső vezetés támogatásának megszerzése.
2. A vállalati célok és feladatok pontos ismerete.
3. A felhasználók folyamatos bevonása, érdekeltté tétele.
4. Meghatározandók a vállalaton belüli információáramlás főbb útjai, csatornái.
5. A VIR hiányosságait pontosan ismerni kell.
6. Az IT-szakértők rendszeres konzultációs bevonása.
7. Kapcsolatteremtés, integráció a groupware és a meglévő alkalmazások között.
8. Hosszú távra való tervezés.
9. Oktatás és tanácsadás biztosítása.
10. Rendszeres felülvizsgálat, lépésenkénti finomítás, bővítés, javítás, módosítás, fejlesztés.



## 3. Felhasználási folyamatok

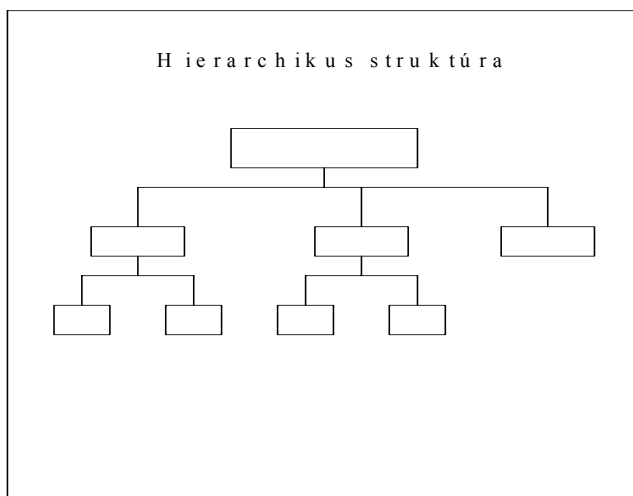
### 3.1. Adatbázis-struktúrák

A következőkben röviden áttekintjük azokat az adatbázis-felépítéseket, amelyek a vezetői információs rendszerekben valósulnak meg. Az adatbázisoknak azért van kitüntetett jelentőségük, mert ezeken hajtódnak végre a vezetők döntési folyamatait kiszolgáló számítások, keresések. Mindehhez a VIR saját adatfelépítésre, saját adatbázisra is támaszkodik, amire azért van szükség, hogy a számításokat minél hatékonyabban lehessen végrehajtani. A vállalati informatikai rendszer adattömegéből ebbe a saját adatbázisba kerül át a releváns vezetői információ.

#### 3.1.1. Hierarchikus: Fa felépítésű

A legkorábbi adatbázis-kezelő rendszerek (Data-Base Management Systems, – DBMS-ek) használták ezt a struktúrát. Ebben a rekordok között fa felépítésű hierarchia érvényesült. Egy rekord fölött csak egyetlen rekord volt, alatta akármennyi. A legfelsőbb adatelem (rekord) az ún. gyökér-elem. Az eléréshez mindig fentről lefelé haladunk a célig. Két rekord között mindig csak egyetlen elérési út létezik, a fa felépítésből adódóan. A rekordok között „egyől sok felé” kapcsolat létezhet. (Angol elnevezéssel: One-to-many relationship.)

Az ilyen jellegű felépítést a 3.1. ábra szemlélteti.



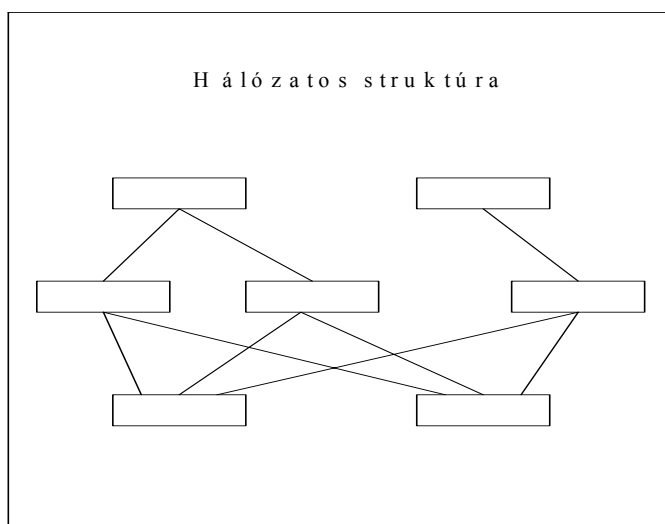
3.1. ábra: Hierarchikus struktúra

### 3.1.2. Hálózatos struktúra

A klasszikus nagygépes rendszereknél terjedt el. Bármelyik eleme kapcsolatban lehet bármelyik másikkal. Az elemek közötti elérés a több lehetséges út egyikén keresztül történik.

Ilyenkor a rekordok között „soktól sok felé” kapcsolat létezik. (Angolul: Many-to-many relationship.)

Az ilyen jellegű felépítést a 3.2. ábra szemlélteti.



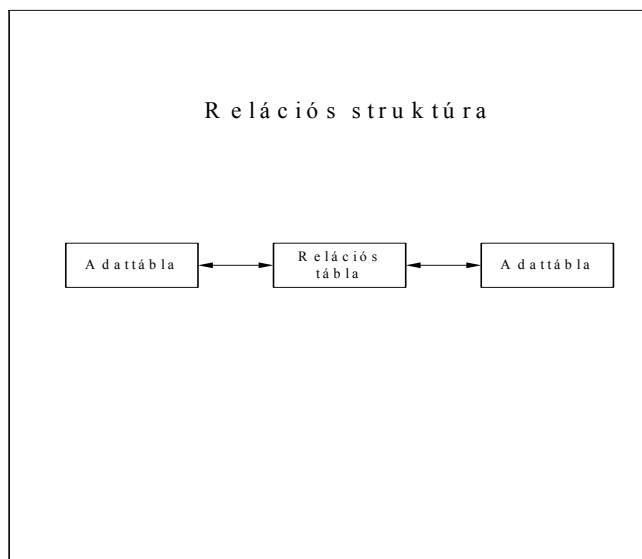
3.2. ábra: Hálózatos struktúra

### 3.1.3. Relációs struktúra (E. F. Codd, USA, 1985.)

Minden adatelem egy-egy táblázatban van ábrázolva. Az adattáblázatok között újabb táblázatok, az ún. relációs táblázatok teremtenek kapcsolatot.

Ez a leginkább elterjedt szervezési megoldás. A modern adatbázis-kezelők is ezen az elven működnek, mint például a legelterjedtebben használt ORACLE.

Az ilyen jellegű felépítést a 3.3. ábra szemlélteti.



3.3. ábra: Relációs struktúra

#### 3.1.4. Többdimenziós struktúra

Az adatbázis elemei egymástól egységes, homogén kezelésmódú függésben állnak. Ebben a szervezésben akármelyik adatdimenzió függhet a többi adatdimenzióhoz tartozó adattól.

Ez a szervezési mód a relációs modellnek egy olyan változata, amely több dimenzióban tárol adatokat és köztük levő relációkat.

Az 1.4. alfejezetben már volt szó arról, hogy a különböző szempontok szerinti on-line analitikus feldolgozás (OLAP) mire alkalmas. Ennek a folyamatnak a hatékony megvalósítását kifejezetten ilyen felépítésű, erre a célra kialakított adatstruktúra tudja lehetővé tenni.

#### 3.1.5. Objektum-orientált struktúra

Modern megoldás. (A mai szoftver-fejlesztések nagy része objektum-orientált szervezésű programokra épül. A Web-fejlesztések alapvető programozási nyelve, a Java is objektum-orientált nyelv.)

Web-alapú alkalmazásoknál használatos, továbbá multimédia jellegű alkalmazásoknál is.

VIR-ek alapszoftverének megvalósításakor, az objektum-orientált adatbázis-felépítés az objektum-orientált szervezésű programoknál vehető figyelembe.

**Objektum:** Műveletekből (operációk), valamint a műveletekhez tartozó saját adatokból (attribútumokból) tevődik össze. A műveletek a hozzájuk tartozó attribútumokon hajtódnak végre. Egy objektum az aktivizálásakor bekerül a memóriába, és a meghívott művelete végrehajtódik, vagyis lefut. Képek, grafika, szöveg, hang kezelése mehet így könnyen (multimédia).

### 3.2. Egy hazai vegyi üzem eladási modellje

A következőkben áttekintést adunk egy hazai vállalat eladási modelljéről, annak adatstruktúrájáról, amire egy tervbe vett VIR épülhetne.

**Üzem:** Rhone-Poulenc Rt, Kazincbarcika. (Ez egy francia gyártó és forgalmazó cég leányvállalata.)

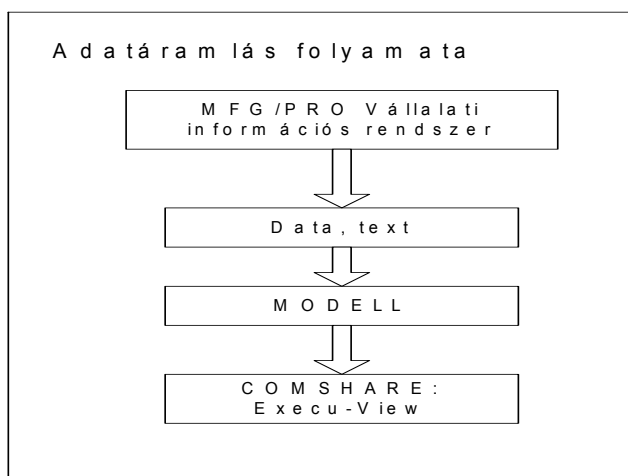
**Termékei:** Ipari és mezőgazdasági vegyszerek, növényvédő szerek.  
Felhasznált *vállalati termelésirányító rendszer:*

#### MFG/PRO.

A tervbe vett 2 Értékesítési (eladási) forgalom figyelése.

**Eszköz:** Comshare rendszer, az Execu-View komponensének bekapcsolásával.

A felsorolt összetevők közötti információs kapcsolatokat a 3.4. ábra mutatja be.



3.4. ábra: Az adatáramlás folyamata

**Felhasználói kör:** Felső vezetés.

**Profillap:** Rhone-Poulenc „SALES” Modell:

**Adatcsoportok és tételszámok:**

- Régiók (10)
- Kereskedők (10)
- Termékek (50)
- Szállítók (200)
- Termék ára (50)
- Eladási volumen (50)

**Forrásadatok:** PROGRESS adatbázis-kezelőtől.

**Konzolidációs adatok:** Milyen mezőkből, milyen nevű és tartalmú mezők keletkeznek, milyen műveletek eredményeként.

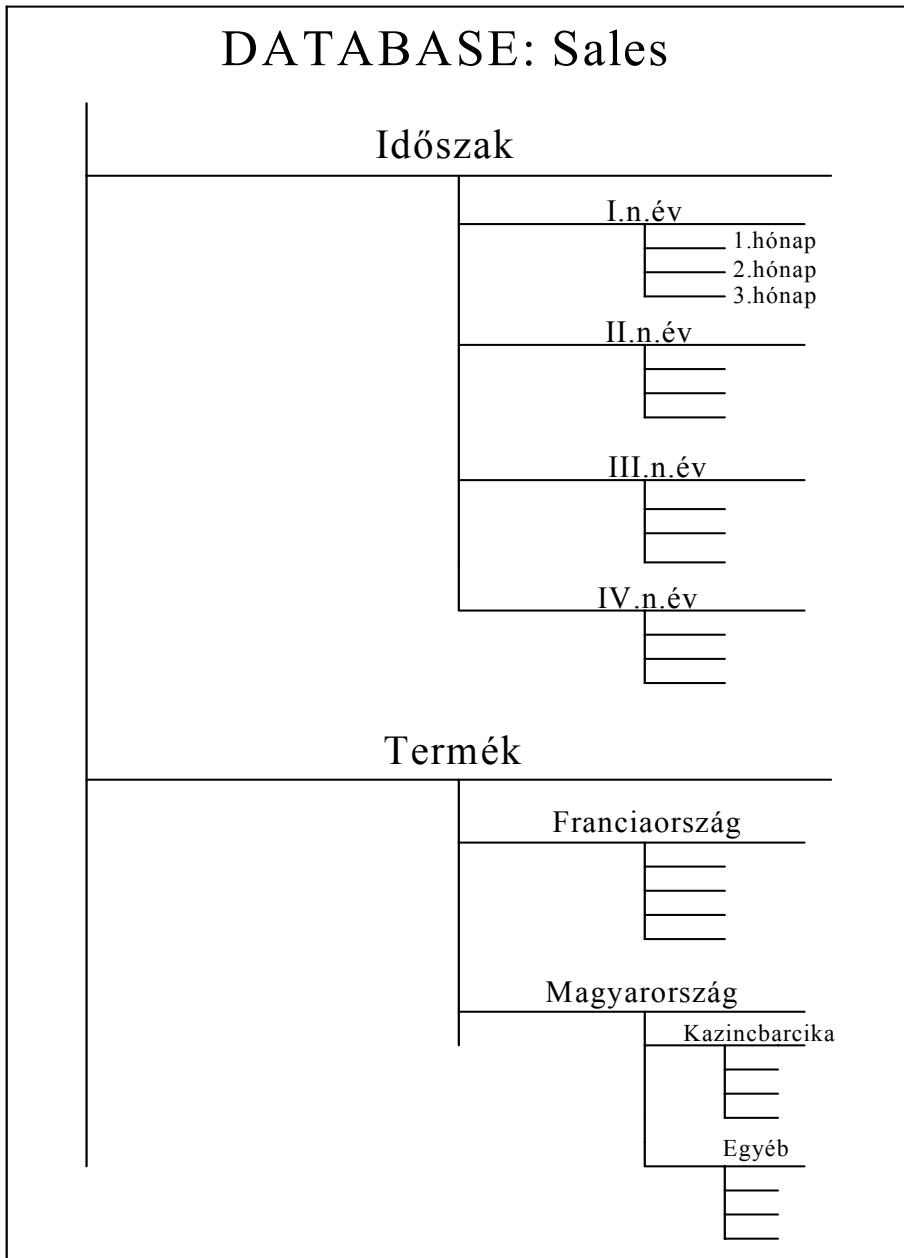
**Modell definíciója:**

**Modell input adatok:** Konzolidált állomány a PC-s alrendszerekből (PROGRESS).

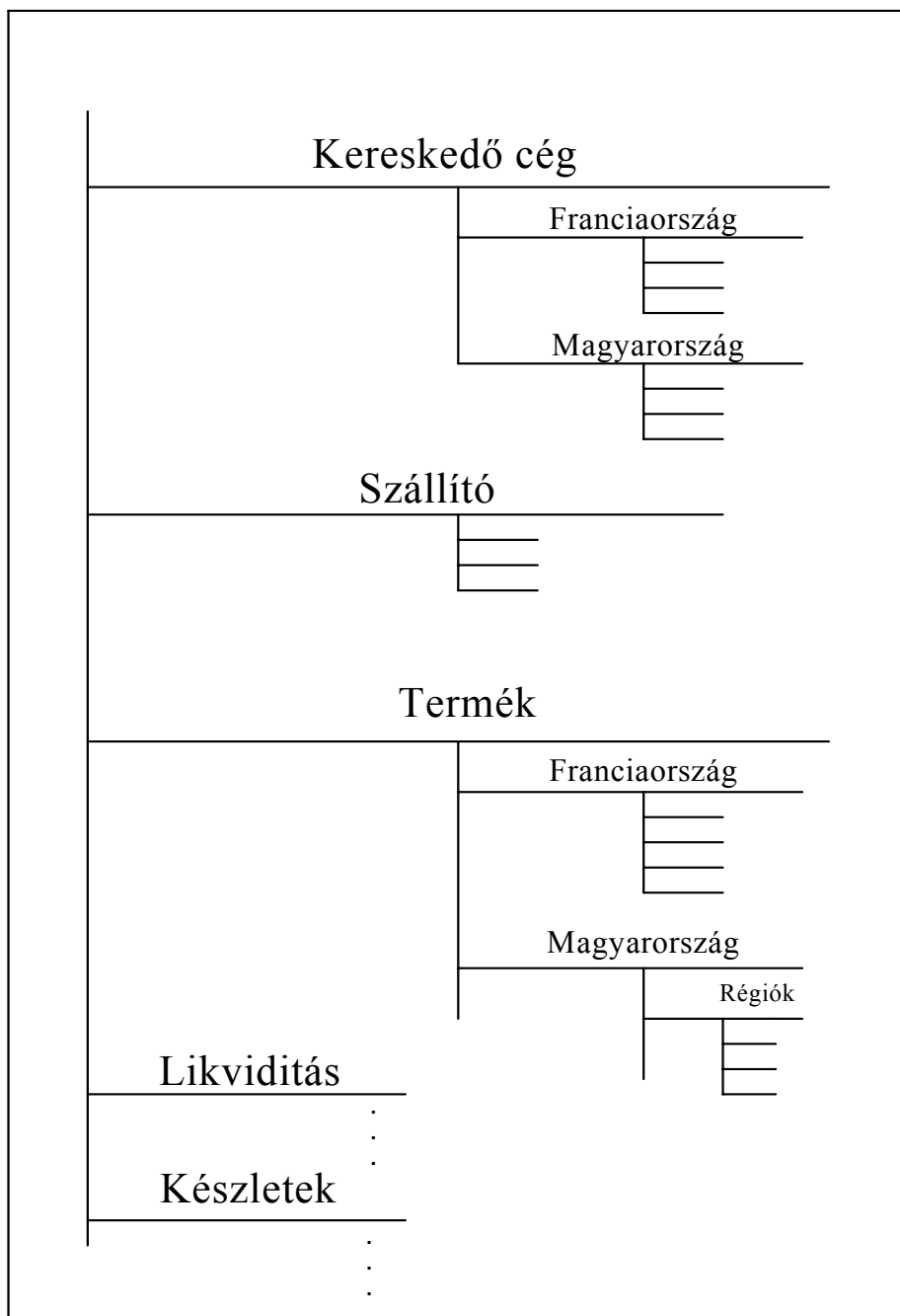
**Dimenziók:**

1. Időszak
2. Termék
3. Változó
4. Régió
5. Kereskedő cég
6. Szállító
7. Eladási volumen
8. Likviditás
9. Készletek

A következőkben az *Eladások (Sales) modell* egy lehetséges többdimenziós kifejtését mutatjuk be (3.5. ábra):



Változó	
	Termék ára
	_____
	_____
	_____
	_____
	_____
	Termék költsége
	_____
	_____
	_____
	_____
	_____
	Valuta átváltási arány
	_____
	_____
	_____
	_____
	_____
Régió	
	Franciaország
	_____
	_____
	_____
	_____
	_____
	Magyarország
	_____
	_____
	_____
	_____
	_____



3.5. ábra: Az Eladások (Sales) adatbázisának modellje



### 3.3. Többdimenziós (multidimenzionális) adatstruktúrák

A multidimenzionális adatbázisok struktúrája a relációs modellnek egy olyan változata, amely több dimenzióban tárol adatokat és köztük levő relációkat. (Egy ilyen modell volt látható a 3.5. ábrán is.)

A tárolás szervezését úgy lehet elképzelni, mint olyan kockákat, amelyek adatokból állnak. És egy ilyen kockán belül újabb kockák helyezkednek el. (Egyik dimenzió függhet a másiktól is.)

Itt a kocka egy oldallapja képvisel egy adatdimenziót. Egy cella a kockában olyan adatot tartalmaz, amely az összes dimenzióra van előállítva: az ilyen cellatartalom neve: *aggregált adat* vagy *konzolidált adat*.

Például, egy cella tartalma lehet:

#### **Eladott mennyiség (termék, régió, eladási csatorna, év, hónap)**

Ez itt öt dimenzió. Vagyis, az eladott mennyiség aggregált értéke a zárójelben felsorolt öt adat aktuális értékétől függ. Az érték itt nyilvánvalóan általánosan, absztrakt módon értendő. Például, a termék most egy adott tárgy, a régió pedig egy földrajzi körzet, megfelelően kódolt formában. A példában szereplő adatviszony tömören így írható le:

**Aggregált adat:** Konkrét adat (5 konkrét dimenzióérték).

### 3.4. Adatkeresési folyamatok a többdimenziós struktúrában

A többdimenziós struktúra térbeli ábrázolására szemléltető példaként a 3.6. ábrán látható kocka szolgál. A kocka alkotóelemei konkrét értékkel rendelkező adatok, amelyek különböző dimenzióértékektől függenek. A dimenziók az elemek oldalán, ill. a kocka oldalán lévő mezőkön vannak feltüntetve.

Két alkotóelem az előoldalán például:

**Eladás** (Videó, Március, Nyugat, Valódi).

Mellette pedig a tervezett eladási érték ugyanerre az esetre:

**Eladás** (Videó, Március, Nyugat, Terv).

Az általános kocka bármelyik építőelemét le tudjuk írni hasonló struktúrában, különböző számú dimenzió felhasználásával.

A sokdimenziós adatbázisban való keresés, vizsgálódás voltaképpen abból áll, hogy különböző beállítások mellett hozunk elő, jelenítünk meg konszolidált értékeket.

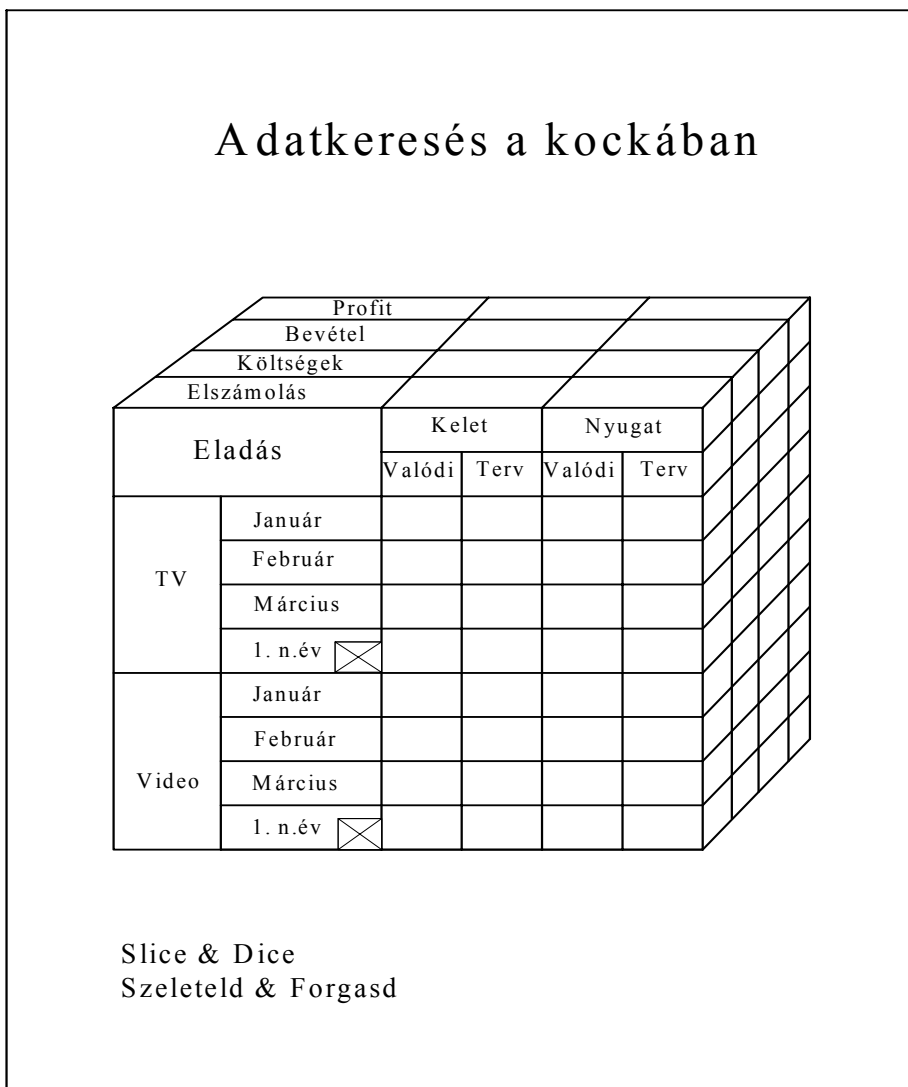
Ez általánosan a következő módon történik:

1. Először rögzítünk egy értékalmazt, amivel tulajdonképpen egy *metszetet* készítünk a kockában. Ezt a műveletet *szeletelésnek* is szokás nevezni (angolul: *slice*).
2. A rögzített értékalmaz mellett, változó adathalmazzal keresünk ki bennünket érdeklő értékeloszlásokat, amit *kockaforgatásnak*, vagy röviden *forgatásnak* nevezünk (angolul: *dice*).

A keresési folyamat a fenti két tevékenység egymás után történő sorozatos ismételtetéséből áll: a szeletelés és forgatás (slice and dice) ismételtetése megy végbe. (Kézzel fogható szemléltető eszköz lehetne ehhez a mindenki által ismert Rubik-kocka.)

Megjegyzés:

Gyakori eset a keresési folyamatban, hogy egy-egy kiválasztott dimenzió mentén végzünk részletesebb feltárást, annak mélyebb vonzatát követjük végig, a keresésben lefelé haladva, mintegy zoomolva. Az ilyen jellegű keresési megközelítést *leásásnak*, vagy *lefúrásnak* (angolban: *drilling*) nevezzük. (Ez a metszetkészítés, és az abból kiinduló forgatás egyszerűbb esete.)



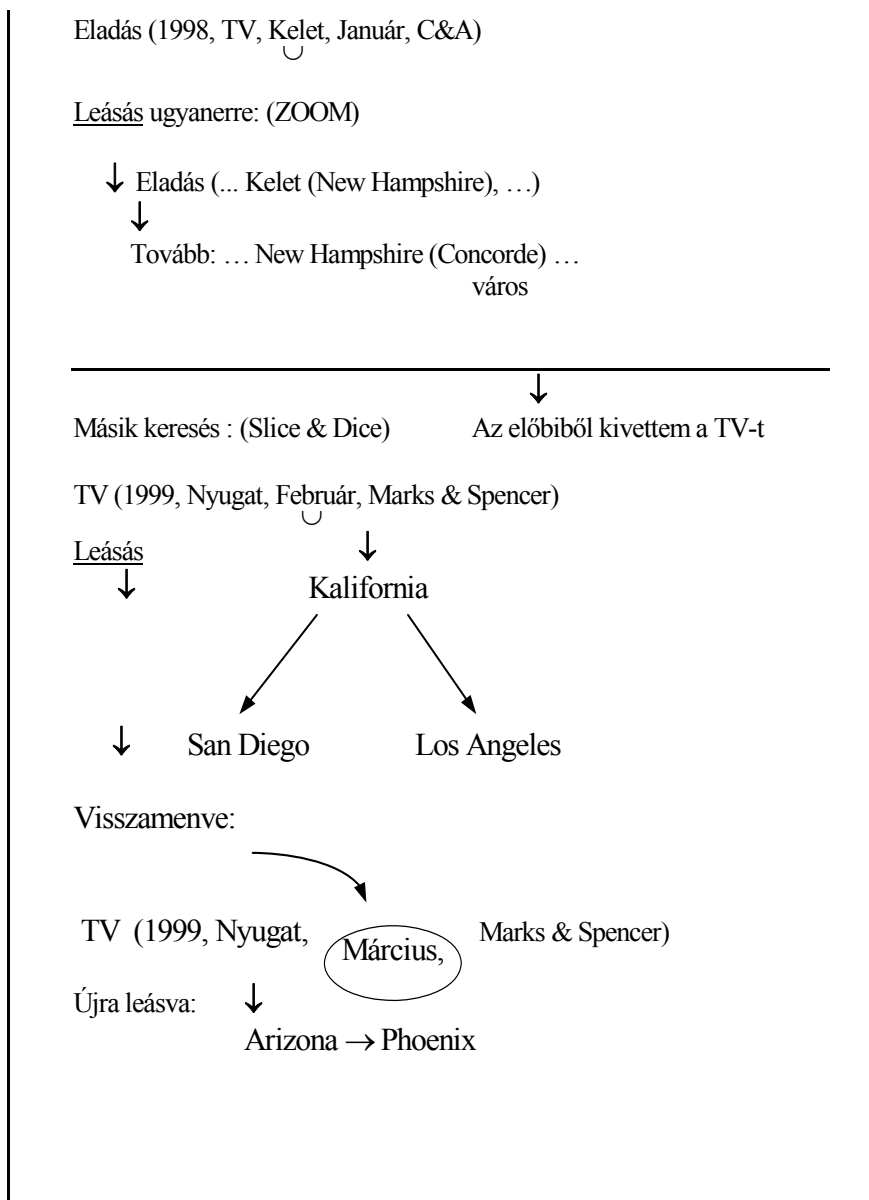
3.6. ábra: Adatkeresés a kockában

### 3.5. Az OLAP végrehajtása

Az előző alpontban leírtak személtetésére szolgálnak a 3.7. és 3.8. ábrán bemutatott folyamatok.

A 3.7. ábra egy-egy dimenzió részletesebben feltárt vizsgálatára mutat példát. Itt jól látható, hogy *leásási lépések* végrehajtására került sor.

A 3.8. ábra egy olyan OLAP-folyamatot tüntet fel, ahol a tervtől való negatív eladási eltérés okának és helyének a keresése megy végbe.



**3.7. ábra:** Keresés és leásás folyamatai

## Egy OLAP és leásás

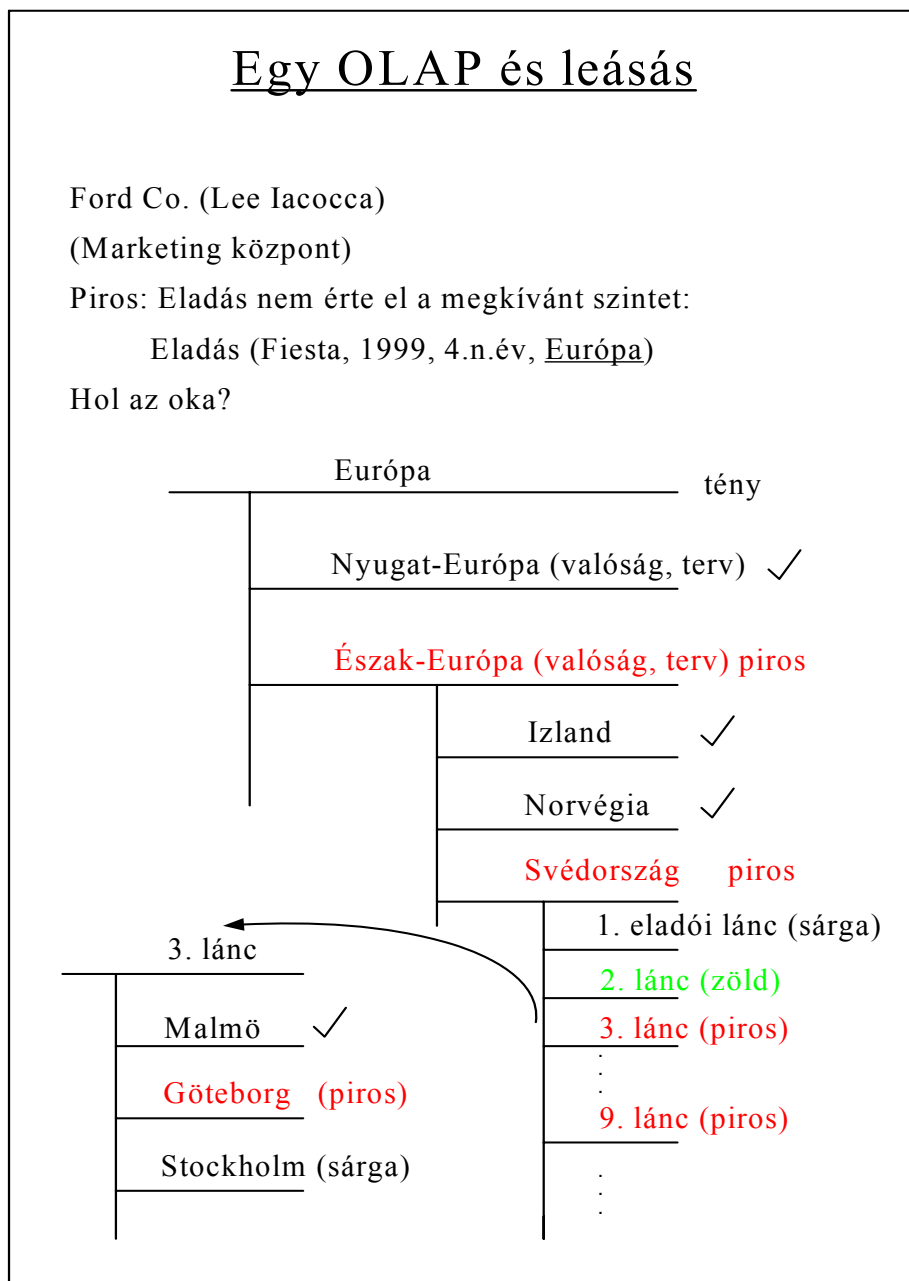
Ford Co. (Lee Iacocca)

(Marketing központ)

Piros: Eladás nem érte el a megkívánt szintet:

Eladás (Fiesta, 1999, 4.n.év, Európa)

Hol az oka?



3.8. ábra: Egy OLAP és leásás folyamatai

Ezután következhet egy átfogóbb vizsgálat a göteborgi eladói lánc tevékenységére vonatkozóan:

Például:

- Az elmúlt 2 év eladási adatai.
- A többi autómárka eladási adatai.
- stb.

Megjegyzések:

- Fontos tudni, hogy a keresés nemcsak **fentről lefelé** haladva végezhető, hanem ahogy a szükség kívánja, **oldalirányban**, vagy pedig egy másik kiválasztott ágon **felfelé** haladva a többdimenziós fastruktúrán.
- Ez a folyamat a jellemző a VIR-ekben a kritikus sikertényező figyelésére, valamint a kiugró eltérések okainak nyomozására.
- Ugyanígy elindítható egy nyomozás a kulcsfontosságú teljesítménymutatók esetében is.
- Pl. sok reklamáció jött be egy termékre. Hol keresendő az oka ?
- Kezdünk lemaradni egy versenytárstól. Mi az oka ?

### 3.6. Szakértői rendszerek használata VIR-ekben

A *Szakértői Rendszer* (SzR) olyan szoftver rendszer, amely olyan emberi szaktudást tartalmaz egy specifikus szakterületen, amit egy felhasználó számára át tud adni.

A szaktudás a rendszer ún. *tudásbázisába* (*Knowledge Base*) van beépítve.

A rendszer a saját ún. *következtetési mechanizmusa* (*inference engine*) segítségével használja fel a tudásbázisban levő információt az aktuálisan fellépő feladatokra, helyzetekre való reagáláshoz.

A következtetési mechanizmus a tudásbázisból való tudáskinyerés segédeszköze.

#### 3.6.1. A tudásbázisok típusai

##### 1. Esetbázisú tudás:

*Esetek* vannak benne, korábban megtörtént esetek, ill. az ezek alapján le-szűrt tapasztalatok összessége. Amire precedens volt már.

##### 2. Keretbázisú tudás:

Hierarchikus kapcsolatban álló blokkok vannak benne. Egy blokk: a *keret*, amelyben összefüggő esetek vannak felhalmozva.

### 3. Objektumalapú tudás:

*Objektumok* hálózata. Az objektumok adatok és az őket feldolgozó műveletek (operációk).

### 4. Szabálybázisú tudás:

A tudás *szabályok* sorozatában van képviselve. Ez a felépítés és szervezés van a leginkább elterjedve.

Egy szabály felépítése:

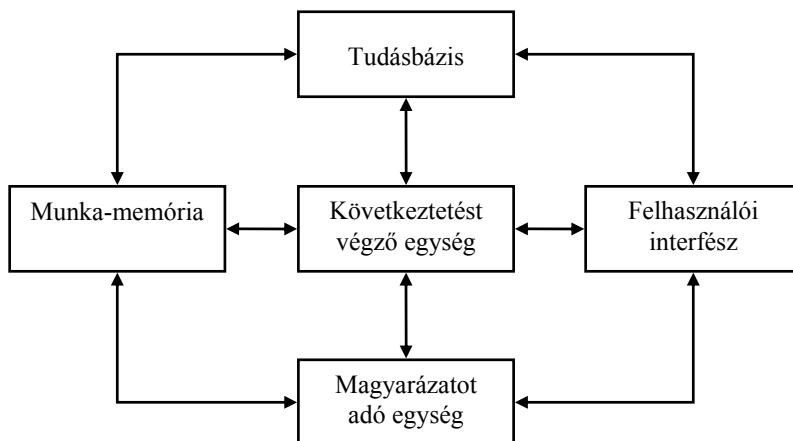
**IF** (feltétel) **THEN** (konklúzió);

Egy feltételben AND/OR/NOT is lehetséges, ill. ezek különböző kombinációja.

Egy tudásbázisban több száz szabály is lehet. A következtetési mechanizmus az aktuális feltételcsoporttal összhangban levő szabályokat válogatja össze, az ellentmondásban levőket ugyancsak elemezve. A végső következtetés az elemzési folyamat befejezése után alakul ki, ami a szakértői rendszer részéről várt megoldást jelenti.

#### 3.6.2. A szabálybázisú rendszerek felépítése

Egy tipikus szabálybázisú rendszer komponenseit és a közöttük levő információk kapcsolatokat a 3.9. ábra mutatja be.



3.9. ábra: Egy szabálybázisú szakértői rendszer felépítése

Az ábrán feltüntetett komponensek funkciója a következő:

- *Tudásbázis*: A szakértői tudás szervezett tárolására szolgál. A szabálybázisú rendszerek esetében a tudásbázis egyrészt a szabályok halmazát, másrészt pedig a kiindulási tényeket, valamint a kiindulási tényekből és a szabályokból levezetett újabb tények halmazát tartalmazza.
- *Következtetést végző egység* (más néven: *következtetési gép*, *-inference engine*): A tudásbázis szabályait és tényeit feldolgozó egység, amely a következtetési folyamatok szisztematikus végrehajtását végzi el. Előreláncolás (előrekövetkeztetés), ill. hátraláncolás, visszaláncolás (hátrakövetkeztetés, visszakövetkeztetés) elvén működhet.
- *Munkamemória*: A feldolgozás segédadatainak és közbülső adatainak átmeneti tárolására szolgál.
- *Magyarázatot adó egység* (*explanation unit*): A felhasználó állandó tájékoztatását végzi, a döntéshozatali folyamatokról ad információt.
- *Felhasználói interfész*: A szakértői rendszer és a felhasználó közötti információs kapcsolatot biztosítja. A rendszer felhasználói vezérlése is ezen az interfészen keresztül valósul meg.

### 3.6.3. Szabálybázisú rendszerek létrehozása

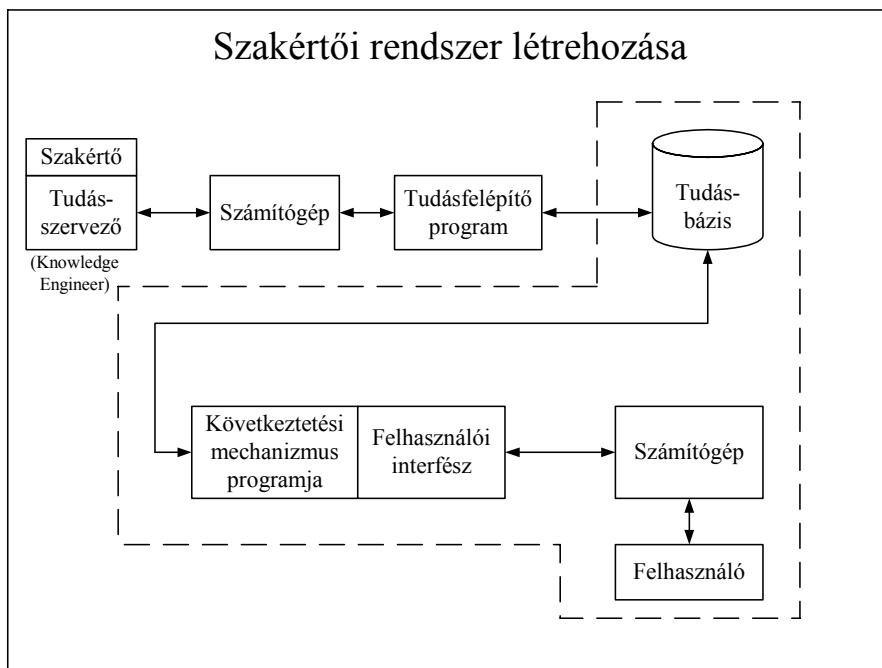
A fejezet hátralevő részében röviden áttekintjük még azt a folyamatot, amelynek során egy szabálybázisú szakértői rendszer számítógépes létrehozása és felhasználása valósul meg. Mindezt a 3.10. ábra szemlélteti.

A létrehozáshoz egyrészt a *szakértőre*, másrészt pedig a *tudásszervező személyre*, vagyis a *tudásmérnökre* van szükség. A tudásmérnök a szakértővel készített interjúk alapján elvégzi a *tudásbázisnak* a szabályokkal való feltöltését. Ehhez az általánosan *tudásfelépítő programnak* nevezhető rendszerkomponenst használja fel.

A létrehozási folyamat alapos teszteléssel, szükség esetén további szak tudás beszerzésével, számos próba futtatással jár. A rendszer elkészülte után kerül sor a valódi üzemeltetésre, amikor is a felhasználó kezelésébe kerül át a rendszer, ahol a vezérlés a *felhasználói interfészen* keresztül valósul meg.

Az üzemeltetésben a *következtetési mechanizmus programja* dolgozza fel a tudásbázis szabályait és tényeit, azzal céllal, hogy egy adott feladatra megoldást adjon a felhasználónak.





**3.10. ábra:** Egy szabálybázisú szakértői rendszer létrehozása és használata

Az ábrán megkülönböztettük a létrehozást és a felhasználást. A két tevékenység azonban nem válik el egymástól egyszer s mindenkorra. A bevitt szabályok módosítása és bővítése a felhasználás során felmerülő későbbi igények alapján végezhető el.

#### 3.6.4. A szakértői rendszerek előnyei és korlátai

Előnyök:

- Egy csoport szakértő tudását gyűjti magába.
- Egyetlen embernél többet is tudhat ilyen módon.
- Gyors és nem fárad el, mindig elérhető, igénybe vehető.
- Ha egy szakértő elmegy a cégtől, a tudása megmarad a cégnél.
- Új emberek tanulhatnak belőle, a tudást hordozó szoftver szétszatható nekik.
- A tudásbázis bővítéséhez olyan szoftver technológia áll rendelkezésre, amely a könnyű és rugalmas változtatást, bővítést teszi lehetővé. (Ún. *Expert System Shell* fejlesztői keretrendszerek.)

Korlátok:

- Nem tanul magától.
- Karbantartása, módosítása, fejlesztése költség.
- A tudásbázis létrehozásához a szakértőt hosszú időre le kell kötni és fizetni.
- Nem képes az új, ismeretlen helyzetekre reagálni. Van benne egy korlátozottság.
- Pénzügyi szabályzatok, politikai kurzus változását nem tudja követni.

### 3.6.5. A Comshare szakértői rendszere

A Comshare vezetői információs rendszer önálló szakértői komponenssel rendelkezik:

#### **Detect & Alert** nevű komponens.

Intelligens SW-komponens, amely automatikusan figyel a többdimenziós adatbázis összes előírt adatértékét. Ha kiugró eltérést, veszélyt észlel (*detektál: detect*), azonnal jelzést ad a felhasználónak róla (*figyelmeztet: alert*), és intézkedési tanácsokkal is ellátja.

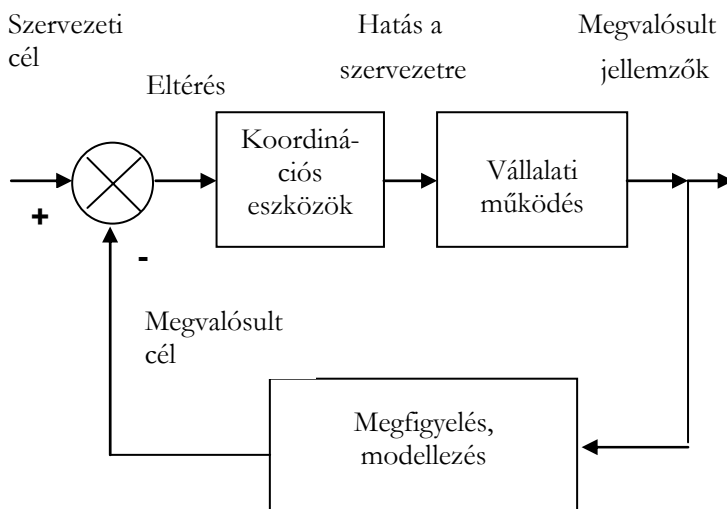
Mit vizsgáljon meg, minek nézzen utána, hogyan próbálja meg kinyomozni a probléma okát, és milyen intézkedésekkel háríthassa el a problémát.

Ez a komponens leásásos elemzéseket is képes végezni az adatbázisban, és azt követően ad tanácsokat.

## 4. Stratégiai menedzsment és informatikai támogatás

### 4.1. Stratégiai folyamatok

A vállalatvezetés rendszerelméleti szempontból hasonló a többi (műszaki, biológiai stb.) vezetési-irányítási helyzethez, itt is érvényesülnie kell a negatív visszacsatolás elvének, amelyet a tantárgy előzményei során már közölt 4.1. ábra mutat be. A vállalat számára ki kell tűzni kívánatosnak tartott, elérendő célokat, és mérni kell, hogy a vállalat tevékenysége milyen mértékben eredményezte a célok elérését. Amennyiben eltérés van a kitűzött és a mért érték között, az eltérés irányával ellentétes irányú hatást kell a vállalatra kifejteni.



4.1. ábra: A vállalatvezetés mint visszacsatolt rendszer

Az, természetesen, nem egyszerű feladat, hogy mit és hogyan kell vagy lehet mérni, és hogy mi az, amire kell vagy lehet célt kitűzni. Nyilvánvaló, hogy csak mérhető érték lehet cél. Ezen túl azonban nagyon sokféle lehetőség van, amelyek azonban nem egyformán célszerűek. Például – lévén, hogy a vállalat szervezet, tehát emberek tevékenységén keresztül működik – az olyan célok a hatékonyak, amelyek teljesítését egy-egy személyen,

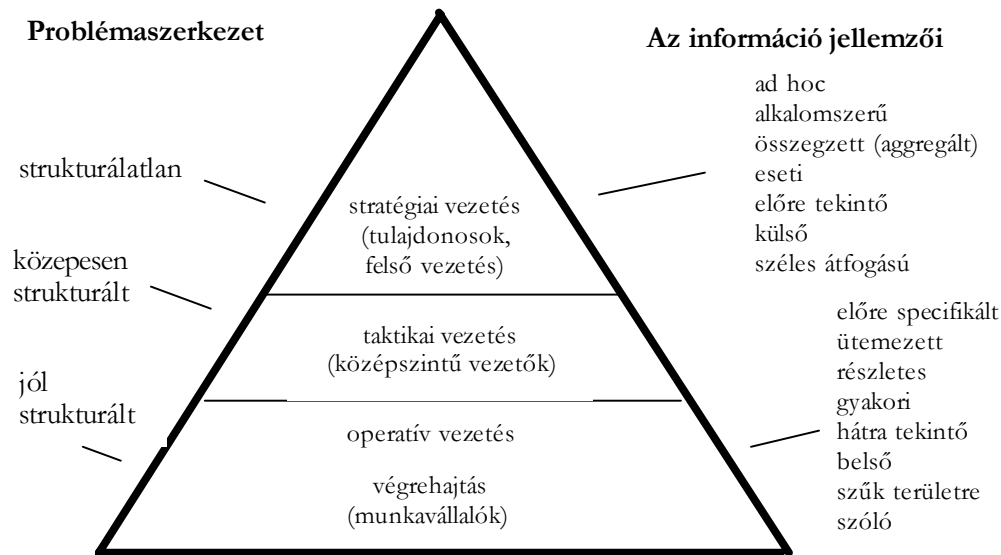
vagy személyek kis csoportján számon lehet kérni azért, mert egyrészt ők döntő befolyással vannak rá, másrészt mert szervezeti intézkedésekkel őket azért felelőssé lehet tenni. A célok helyes kitűzése tehát művészet, aminek az eredményét csak a tapasztalat tudja, utólag, igazolni.

A vezetésben felmerülő kérdéseknek, döntési helyzeteknek és a rendelkezésre álló információ jellegének szerkezete erősen függ a vezetőnek a vállalatirányításban elfoglalt helyétől, pontosabban szintjétől a vállalatirányítási piramisban. A 4.2. ábra megismétli a tantárgy egyik előzményében már közölt szerkezeti ábrát.

A piramis alján, az operatív (műveleti) vezetésben az eldöntendő kérdések egyértelműek, a válaszhoz rendelkezésre áll a jól strukturált információ. Itt a döntések olyan típusúak, hogy például mikor, mennyit és melyik szállítótól kell venni egy alapanyagból vagy alkatrészből.

Ezzel szemben a piramis tetején, ahol a stratégiai döntések történnek, a kérdések maguk sem egyértelműek, néha az a kérdés, hogy mit is kérdezzünk, ennek megfelelően a rendelkezésre álló információ sem jól strukturált, sem homogén, gyakran még az is kérdéses, hogyan lehet hozzájutni? Olyan kérdések merülhetnek fel, hogy egyáltalán van-e szükség beavatkozásra, tehát hogy van-e beavatkozást igénylő eltérés a célok és eredmények között? Ha igen, hol és milyen jellegű legyen a beavatkozás? Sőt, ezen a szinten merül fel feladatként a stratégiai céloknak a kitűzése.

A két szélsőség között folyamatos az átmenet mind a döntések mind az információ jellegében.



4.2. ábra: A döntéshozatali hierarchia és információigénye

### **Példa a strukturálatlan, stratégiai döntésre**

Az 1980-as évek első felében a számítógép-memória integrált áramkörök piacán addig piacvezető Intelnél azt vették észre, hogy az előző évhez képest az ebből az üzletágból származó bevételük stagnál. Szó sem volt még visszaesésről, de elfogyott az addigi lendületes gyarapodás. Ez nem látszott ugyan drámainak, mégis gyanút fogtak, és utána néztek, részletes piaci információt gyűjtöttek, és megtudták, hogy több japán gyártó is fejlődik lendületesen ebben a piac-szegmentumban. Ez a hír már drámaibbá tette a másikat. Megfontolták, mit jelentene felvenni a kesztyűt a japánokkal, mit jelentene belemenni egy éles költségcsökkentési hajszába nekik, akik nem ebben, hanem a technikai és technológiai innovációban voltak igazán jók. 1985-ben végül úgy döntöttek, hogy kiszállnak ebből a szegmentumból, és az addig a forgalmuknak csupán 15%-át kitevő mikroprocesszorokra összpontosítanak, hiszen ahhoz értenek – és ott még nincsen ilyen konkurencia, és valószínűleg később lesz csak, hiszen ahhoz nem elég a meglehetősen általános elveken alapuló, és ezért viszonylag könnyebben elérhető gyártási technológia, hanem nagyarányú szoftver know-how is kell, amit sokkal nehezebb megszerezni. Döntésük helyességét bizonyítja, hogy ezen a területen azóta egyfolytában piacvezetők, olyannyira, hogy 2006-ban még az Apple is úgy döntött, hogy, korábbi stratégiai irányvonalával szakítva, Intel processzorokat épít be a MacIntoshba. És ezt az ügyes piaci politikán kívül azzal érték el, hogy technikailag mindig a legjobbak között voltak, így nem volt szükségük marketing-ügyeskedéssel kiépített monopolhelyzetre.

A vezetői információs rendszerektől elvárják, hogy nemcsak a piramis alján nyújtsanak támaszt, megbízható és jól szervezett adatszolgáltatással, az adatok vezető-barát elérésével és megjelenítésével, hanem hogy a piramis tetején felmerülő problémák megoldásában, sőt, már a megfogalmazásában is segítsenek. Ebből látható, hogy ez a szakterület csak a műveleti jellegű információs rendszerek fejlődésének meghatározott szintjén indulhatott igazán fejlődésnek, és hogy számos olyan jellegzetessége van, amely a vállalati műveleteket támogató rendszereknek nem volt. Az is következik ebből, hogy ezen a területen még ma is nagyobb a sokféleség, ma is gyakoribb a valamilyen értelemben specializált rendszer, és kevésbé vannak szabványszerűen elfogadott módszerek, megoldások, mint a műveleteket támogató rendszerek esetében.

## 4.2. A stratégiai menedzsmentben használt módszerek

A vezetést támogató és gyakorlatilag minden nagyobb vállalatnál használt eszköz a **kontrolling**. Erről viszonylag részletesen volt szó a Vállalati információs rendszerek II. tantárgyban. Ez a pénzügyi menedzsment eszköze, amely biztosítja a rövid távú (első fokon éves) üzleti terv készítésétől, ellenőrzésétől és megvalósításától kezdve a stratégiai pénzügyi tervezés megvalósításáig a vezetés támogatását. Ez áll a legközelebb a műveleti támogató rendszerekhez, azon belül is a számvitelhez, számos gyártó (többek között az SAP is) azzal integráltan szállítja.

A pénzügyi célok megvalósítása tágabb értelemben vett feltételeinek is tekinthető marketing-, termelési, logisztikai, humán erőforrás és végül az informatikai stratégia megvalósítását már más eszközökkel lehet és kell támogatni. Ezeknek a területeknek is vannak ma többé-kevésbé kiforrott, részben a számítógépes támogatás előtt már alakulóban lévő módszerei és eszközei, ám a számítógép itt is az új lehetőségek széles skáláját nyitotta meg.

A stratégiai elemzés, ezen belül a környezet elemzése, a versenypozíció, az erőforrások, a kompetenciák, stratégiai képességek elemzése, az értéklánc-elemzés, a költség-hatékonyság elemzése, az alkalmas mintához hasonlítás (benchmarking), a pénzügyi és egyensúlyi elemzés, a SWOT-elemzés, a kritikus siker-tényezők definiálása és alakulásuk elemzése mind-mind eszköz a stratégiai vállalatirányításban, és mindegyikhez található támogató rendszerek. Egy-egy nagyobb rendszer többnyire számos efféle eszközt támogat, és vannak specifikus területekre specializálódottak is.

Hasonlóképpen a tevékenységi területenként lehetséges stratégiai vezetési eszközöket, módszereket, tehát a marketing, a termék/piaci, a termelési, a pénzügyi, a logisztikai, a humán erőforrás és az informatikai stratégia készítését és megvalósítását is támogatják az univerzális rendszerek megfelelő moduljai vagy specifikus rendszerek.

A következő alfejezetben rövid áttekintést adunk azokról az eszközökről, amelyek éppen azért szükségesek és hasznosak, mert a legmagasabb szintű támogatás döntési és információ-szerkezete strukturálatlan, és az ezen a szinten hozott döntésekben lényegesen nagyobb a döntéshozó intuitív hozzájárulása, ezért itt másként, és sokkal nagyobb mértékben ját-

szanak szerepet az ember információ-kezelési és feldolgozási adottságai, mint a vállalati információs rendszerekben általában.

### 4.3. A menedzser-barát informatika

A számítógépes információs rendszerek hihetetlen hatékonysággal tudnak nagy tömegű adatból hasonlóan nagytömegű újabb adatot gyártani. Az ember azonban nem (nemcsak, és elsősorban nem) ezt igényli. Az ember – és a vállalatvezető is az – emberi formájú információra vágyik. Ezért a vezetői információs rendszerekben nagyon fontos, a műveleteket támogató számítógépes információs rendszerekhez képest sokkal fontosabb minden eszköz, amely az információ felhasználó-barát elérését segíti elő.

Ezek egyik csoportja az adatok keresését, másik csoportja azok feldolgozását, a harmadik pedig közvetlenül a tállását végzi, bár a három gyakran nem igazán választható szét. Jóllehet ma már a vezetői munkakörökben is többnyire a számítógéppel ismerős emberek ülnek, mégis, az ő munkájuk hatékonysága is erősen függ attól, hogy a gép mennyire alkalmazkodik az emberi pszichéhez. Mert a vezető csak igen korlátozottan ér rá arra, hogy ő alkalmazkodjék a géphez. Megszokja a gép kezelését, a billentyűzetet, a képernyőt, a betűtípusokat, az egeret, de azt elvárja, hogy a gép következetesen – sőt, mindig egyformán – viselkedjék, és hogy ő maga a saját, egyéni, vezetői logikája szerint tudjon eljárni, ebben a gép ne akadályozza, hanem segítse. Bár számos eszköz, módszer segíti, hogy ember és gép egymásra találjon, itt csak néhány, a vezetői rendszerekben különösen fontosról fogunk szólni.

Egyik ilyen, az imént említett első csoportba (keresés) tartozó eszköz az **OLAP-rendszerekben** megvalósított sokféle keresési, forgatási, leásási eljárás. Ezek elsődleges célja az, hogyha a vezetőnek eszébe jut hirtelen, hogy az adatbázis egészen más részében és máshogyan tárolt adatok valamilyen újabb összesítésére vágyik, azt azonnal megkaphassa. Azonnal, hogy ne zökkenjen ki az eredeti gondolatmenetéből azzal, hogy hosszasan keresgél, sőt, magát a formulát keresgéli, azt, amivel eljuthat a kívánt terület adataihoz. Az OLAP-rendszereket önálló fejezetben fogjuk tárgyalni.

Másik, az iménti harmadik, tállási csoportba tartozó eszköze az ember-barátságának a **vizualizálás**. Itt olyasmikről van elsősorban szó, mint a diagramok, amelyek sok számot ábrázolnak jól látható formában, a numerikus megadás helyett, vagy viszonyokat (relációkat) ábrázolnak a körülményes (bár esetleg szabatos) szöveges leírás helyett. Vizualizálás a színnel való kiemelés, kódolás is. Ennek a háttérében az van, hogy az embernek



vele születetten rendkívül fejlett a látási (vizuális) ingereket feldolgozó idegrendszeri apparátusa, a teljes, érzékszervekhez kötött agykéreg több mint 90%-a a szemmel kapcsolatos, 5%-a a füllel, és a maradék a többi érzékszervvel (szaglás, tapintás stb.). Ráadásul a vizuális ingereket sok önműködő, tehát gondolkodást, tudatos tevékenységet nem igénylő idegrendszeri művelet elemzi, szűri, létezik az ún. vizuális gondolkodás, vizuális problémamegoldás. Ezért ez a rendszer kiemelkedően hatékony az ember esetében. Nem használni tetemes elmaradt haszonnal jár, ami egy vezető esetében türelmetlenséget szül.

A műszaki életben nagyon régen (már a régi görögök is) használták a rajzot, eleinte spontán, később szabványosított formában, tervezési fázisban. Az üzemi-üzemeltetési fázisban, a folyamatirányításban a munkavégző személyek szinte már a kezdet kezdetén igényeltek és kaptak vizualizáló eszközöket. Ugyanis a villamos kapcsolótáblákat nagyon korán elkezdték úgy kialakítani, hogy színes lámpákkal megjelenítették a kapcsoló eszközök állapotát, mutatós műszerekkel a mért mennyiségek (áram, feszültség, teljesítmény stb.) értékét, sőt, úgy helyezték el a táblán a kijelző- és kezelőszerveket, hogy az elrendezésük a folyamatbeli összefüggésüket is kifejezze, és ezt még ábrákkal (festett szimbólumokkal) is támogatták.

Így alakultak ki az ún. sématablák, amelyeket legelőször a villamos erőművi hálózatokban használtak, de hamarosan nemcsak ott, hanem más, elsősorban a vezetékes és folytonos technológiákban is, mint például a vegyiparban vagy a közműveknél (víz, gáz, csatorna). Azért először itt, mert a technológia mintegy sugallja a megjelenítés módját, a célszerű, a kezelő munkáját, ítéletalkotását, döntését támogató ábra szerkezetét. Másrészt egy nagyobb rendszer áttekintése gyakorlatilag lehetetlen is egy ezzel a feladattal megbízott személy számára ilyen segítség nélkül. Ezek a táblák lényegében megismételték az elvi kapcsolási rajzok szerkezetét. Az első számítógéppel támogatott efféle rendszerekben a katódsugárcsöves megjelenítő eszközökön ezeket a sémákat ismételték meg, csak immár szoftver állította össze, és mozgatta az ábra elemeit.

A közgazdaságtanban is hamar rájöttek, hogy ugyan a számvitel (könyvelés) nem vizualizálható, de már a többi közgazdasági tevékenység és viszony leírását is hatékonyabbá teszi az ábra. Ma már szinte nincsen olyan szellemi tevékenység, ahol ne használnának – legalábbis segédeszközként – ábrákat. Legújában a tőzsdei árfolyam-mozgások vizualizálása okoz izgalmat a szabályozó hatóságok és spekulánsok körében egyidejűleg. Hamarosan kitérünk erre a szakterületre, egy önálló alfejezetben.

A feldolgozási csoportba tartozó eszközök a mesterséges intelligencia különböző megközelítései, olyan módszerek, amelyek – többnyire a felhasználó interaktív közreműködésével – az adatokban rejlő információk kiásását, az adatokból a teendők meghatározását teszik lehetővé. Efféle az **adatbányászás**, amivel szintén külön fejezetben foglalkozunk. Efféle munkát végeznek a nemrég említett **szakértői rendszerek** is.

A feldolgozási csoportba tartoznak továbbá a **mutatószámok**. Ezek is használatosak voltak már a számítógép feltalálása előtt is, de itt megújulva, sokkal szélesebb és változatosabb lett a felhasználásuk. A háttérükben az a lélektani tény áll, hogy az emberi agy nem képes egyidejűleg sok dolgot kezelni. Természetesen egyénenként nagy a szórás, és megfelelő edzéssel sokat lehet javítani ezen a képességen is, de tény, hogy a legtöbb ember egyszerre legfeljebb 7 különböző dolgot tud hatékonyan kezelni. A nagy adathalmazokat tehát strukturálni kell (és lehetőleg vizualizálni a struktúrát), és az így létrejövő szerkezetekben, egyenként 5-7 elemnek szabad lenni. Egy fa-szerkezetű halmaz (ábrázolva: gráf) esetén az egyes csomópontokból legfeljebb 7 ágnak szabad indulni, illetve az egyedek felől indulva, egy csomópontban legfeljebb 7 egyedből jövő ágnak szabad egyesülni. Az ezzel a rendszerezéssel létrehozott magasabb rendű egyedeket (entitásokat) pedig egy vagy néhány saját adattal jellemezzük. Ezek a mutatószámok.

A mutatószám tehát az egyedi adatok sűrítése, általában több, néha nagyon sok egyedi adatból állítják őket elő. Ilyen eset, amikor mutatószám keletkezik a sok adat valamilyen feldolgozásából, például összeg- vagy átlagképzésből.

Vegyük példaként a sajtóban unalomig emlegetett bruttó nemzeti összterméket (GDP)! Ebben benne van a vonatkozó ország vagy térség és időszak összes mért, értéknövelő ügylete. Ez egyúttal jó példa is arra, hogy milyen nehéz a célnak megfelelő mutatószámok megalkotása, kiválasztása. A GDP például majdnem jól mutatja, hogy mennyi munkát, erőforrást használtak fel, de azt nem, hogy mekkora volt a fejlődés. Gondoljanak arra, hogyha valamit rosszul csináltak meg, le kellett bontani, és újra építeni, az kétszer szerepel a GDP-ben, plusz még a bontási ráfordítások. Tehát éppen arra, a térségek fejlettségének összehasonlítására, nem alkalmas, amire a legtöbbször használják, hiszen az ostobaság és pazarlás is növeli az értékét. A ráfordítást pedig azért nem méri helyesen, mert a GDP sem méri az olyan értéknövelő tevékenységeket, mint például az anya által végzett gyereknevelést – holott lehet, hogy társadalmilag ez a legfontosabb

mind között, és az is lehet, hogyha szigorúan elemezzük, közgazdaságilag is, kiderül, hogy egy jó nagyságrenddel nagyobb ezeknek a nem mért értékeknek az összege, mint a mértéké. De ez nem a GDP hibája, hanem azoké, akik arra használják, amire nem való.

A mutatók egy része valamilyen mértékegységgel is ellátott, abszolút szám, a GDP mértékegysége például valamilyen pénznem. A másik részük a relatív mutató, ezt gyakran százalékosan adják meg, például a külkereskedelmi deficitet a GDP százalékában. Erről a témáról egy hamarosan következő alfejezetben szólnunk.

#### **4.4. Algoritmusok adat-elérési, szűrési, elemzési célra**

Ez az első, ami általában, nekünk, informatikusoknak eszünkbe jut a vezetési tevékenység számítógépes támogatása esetében: az ötletes, szabadalommal is védhető eljárás, ami lehetővé teszi, hogy a felhasználó a hatalmas mennyiségű adatból könnyen és gyorsan elérje éppen azt, amire szüksége van. A vezetői információs rendszerek többsége is úgy született eredetileg, hogy az adatbázis-kezelés specialistái kifejlesztettek egy hatékony tárolási és elérési módszert, és azt körberakták felhasználói és interfész modulokkal.

Az algoritmikus rendszerek alapvetően kétfélék:

- adat-elérést segítő és
- adatfeldolgozó,

algoritmusok. Mindkettőben szerepet játszanak

- kereső és
- szűrő

algoritmusok.

Az elérési algoritmusokhoz tartoznak maguk az adatbázis szervezési megoldások, és a hozzájuk tartozó keresők. A feldolgozók nagyon sokfélék, mert az adathalmaz jellegének és a feldolgozás céljának megfelelően más-más megoldásokra van szükség. Az adatbázisokkal szaktárgyakban már találkoztak. Ebben a tárgyban az algoritmikus alapon felépített rendszerek közül a továbbiakban két csoportról fogunk külön szólni, amelyek már a vezetői munka célszerűsége szerint szerveződtek rendszerbe:

- a kereső-elemző OLAP (on line analytical processing) és
- az adathalmazok rejtett tulajdonságainak feltárásában segítő adatbányászó (data mining)

rendszerrel, mindkettővel egy-egy későbbi fejezetben. Korábban említettünk még egyet, az emberi tudást alkalmazó szakértői rendszereket. A könyv végén lévő fejezetben bemutatott Comshare rendszernek a Detect & Alert komponense például szakértői rendszert használ annak meghatározására, hogy az adatok milyen állása mellett kell a felhasználót riasztani.

## 4.5. Mutatók, mutatószámok

### 4.5.1. A mutatószámok szerepe

Amint korábban megállapítottuk, a nagy adathalmazoknak, így a számítógépes eszközzel kezelteknek is, az ember számára jól kezelhetővé tételéhez hasznos eszköz a *mutatószámok* használata. A használatuk jóval a számítógép előtti időkre tekint vissza, a fejlett kapitalista gazdaságban mindenütt használták őket, és a részvénytőzsdére bevezetett vállalatoknak elő is írták – és ma is előírják – bizonyos mutatók nyilvánosságra hozatalát. A számítógépes információs rendszerek azonban korábban nem várt távlatokat nyitottak a felhasználásuk számára. Ezek a rendszerek ugyanis egyrészt képesek nagyszámú adatból, bonyolult összefüggések alapján, nagyszámú mutató kiszámítására, másrészt beprogramozhatók arra, hogy adatokat, közöttük akár mutatókat, figyeljenek, és azok változása alapján figyelmeztető jelzéseket küldjenek a felhasználóknak.

Ezzel emberi munkát takarítanak meg, a vezetőnek nem kell magának megjegyezni a határértékeket és figyelni azok túllépését, elvégzi ezt is a gép. Másrészt a gép szinte korlátlan számú mutatóval tudja ezt egyidejűleg elvégezni, az ember viszont, mint ezt korábban kifejtettük, nem. Ezzel túllépnek az ember helyettesítésén, igazi értéknövelő támogatást nyújtanak. Az ember számára már áttekinthetetlen adathalmazon is eredményes tud lenni a gép, csak pontosan elő kell neki írni, mit tegyen. Végül a felhasználó számára egy már szűrt, általa is áttekinthető mennyiségű információt talál. Az ilyen jellegű figyelést, és a kiugró eltérésekből származó riasztást például a Comshare Detect & Alert komponense szakértői rendszer felhasználásával valósítja meg.

A mutatószám tehát az egyedi adatok sűrítése, általában több, néha nagyon sok egyedi adatból állítják őket elő, gyakran összetett számítással.

A mutatók egy része valamilyen mértékegységgel is ellátott, abszolút szám, a GDP mértékegysége például valamilyen pénznem. Az efféle mutatók lehetnek egyrészt:

- eredeti adatok, például árak, darabszámok, egyéb mennyiségek (pl. eladott folyadék mennyisége, literben vagy köbméterben),
- származtatott adatok, az eredeti adatokból összegzéssel, különbségképzéssel stb. származtatva.

A másik részük a relatív mutató, ez lehet:

- százalékos adat, tehát azonos mértékegységű adatok aránya, például a külkereskedelmi deficit a GDP százalékában,
- egyéb viszonyszám, például adott időszakra vagy egy főre eső forgalom, darab, árbevétel.

A mutatók használata kockázatokat is rejt, ami egyébként mindenfajta sűrités esetében fennáll. A mutató ugyanis egy számítás végeredménye, még ha az egyszerű összeadás is, és ezáltal már nem látszanak benne az egyes összetevők által hordozott információk. Például ha egy mutatónak, amelyről tudjuk, hogy két érték összege, az értéke 9, ez létrejöhet 1+8-ként és 4+5-ként is! Ezen – persze elsősorban nagyobb adathalmazok esetében – több, egymás információját kiegészítő mutató használatával lehet segíteni.

A statisztikában, ahol rengeteg mutatót használnak, amelyeket azonban valószínűségelméleti alapon, tudományosan definiálnak, például gyakran használnak a sokaságot általában vagy az egyedek eltéréseit jellemző további mutatókat. Például azt mondják, hogy a sokaság (az adatoké) jó közelítéssel normális eloszlású,  $K$  középértékkel és  $S$  szórással (egy elméleti alapokon nyugvó értelmes kijelentés az egyedi adatok egymáshoz képesti nagyságáról és két mutató). Ez így együtt már sokkal többet mond (volta-képpen bármilyen statisztikai célra ennyi elegendő is), mint az egyetlen középérték (magyarul átlag). Amikor nem lehet ilyen szép, tudományos kijelentést tenni, hogy „normális eloszlású”, akkor használhatók egyéb mutatók, például a medián (az értékek halmazának közepe), vagy fel lehet osztani a sokaságot decilisekre (10 sávra) vagy kvartilisekre (4 sávra), és megmondani az azokban lévő adatok számát.

Mindezek a módszerek sokat segítenek abban, hogy a sűrités okozta információvesztés minél kisebb legyen, de teljesen elkerülni nem lehet. Ha nincsen semmilyen elméleti alapunk, hogy az adatok eloszlását megjósol-

juk, célravezető lehet az adathalmaz kluszter-elemzése, vagy más, adatbányászási technika alkalmazása a tömörítés és az értékes információ dilemmája megoldását szolgáló kompromisszum érdekében. Ezzel a témával később önálló fejezetben foglalkozunk.

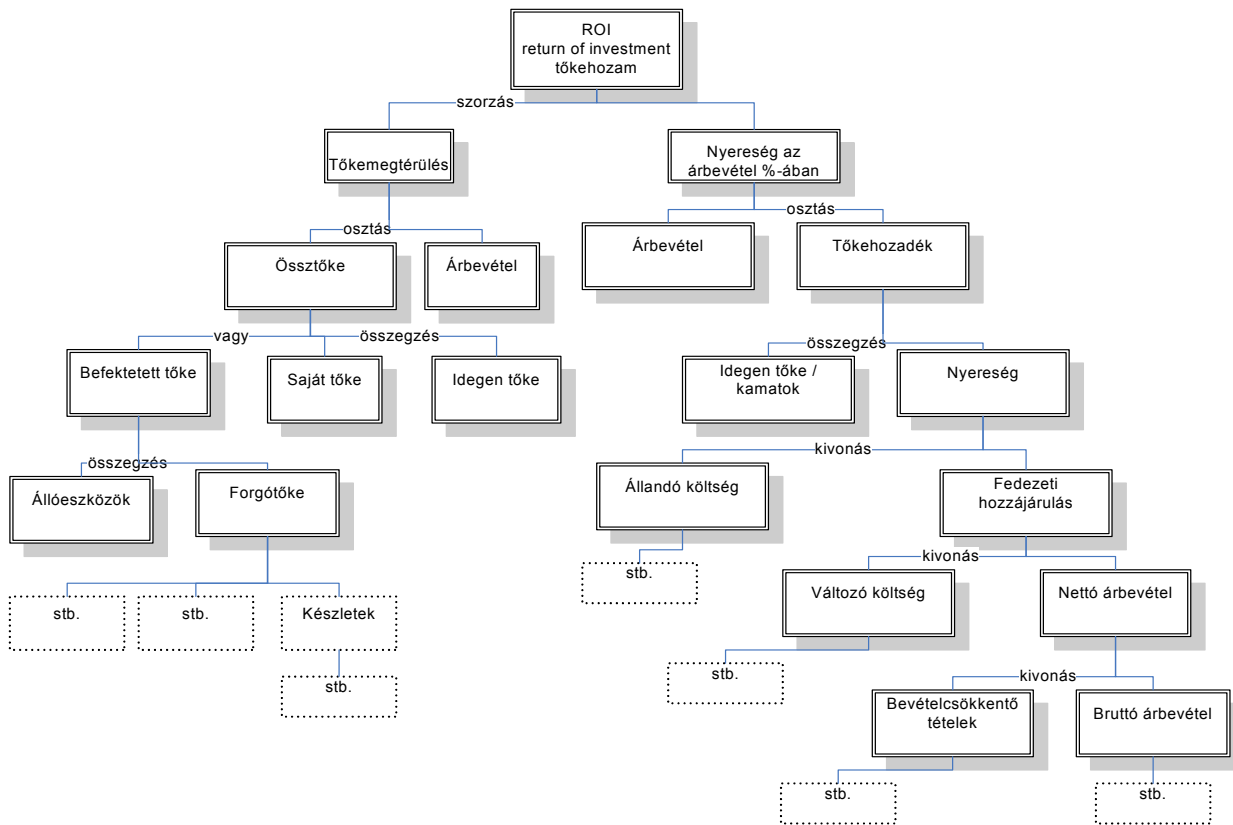
A dolog másik oldala viszont, hogy a mutatót hogyan definiáljuk, hogy minél inkább valami olyat „mutasson”, aminek számunkra jelentősége van, tehát valami fontosat. Ez egyáltalán nem egyszerű feladat, és többnyire csak a gyakorlat tud a lehetséges ötletek között választani. Annyi eleve világos, hogy sohasem szabad egyedí, elkülönült mutatókat definiálni, hanem mindig rendszert kell, egymással összefüggő, egymás információját adott cél érdekében kiegészítő mutatókból álló rendszert. Ehhez először a célt kell világosan, egyértelműen megfogalmazni, abból majd adódik a legjellemzőbb és legfontosabb adatok, illetve a belőlük képezhető mutatók köre. Még ezután is sokféle lehetőség marad, és bizonyos fokig ízlés dolga. A kockázatok csökkentésére itt is – mint azt annak idején, a vállalati információs rendszerek fejlesztésénél láttuk – célszerű a már bevált mintákat követni, használni, és csak elegendő saját tapasztalat birtokában hozzálatni a módosításhoz, továbbfejlesztéshez.

#### **4.5.2. Mutatószámok és a vezetési tevékenység**

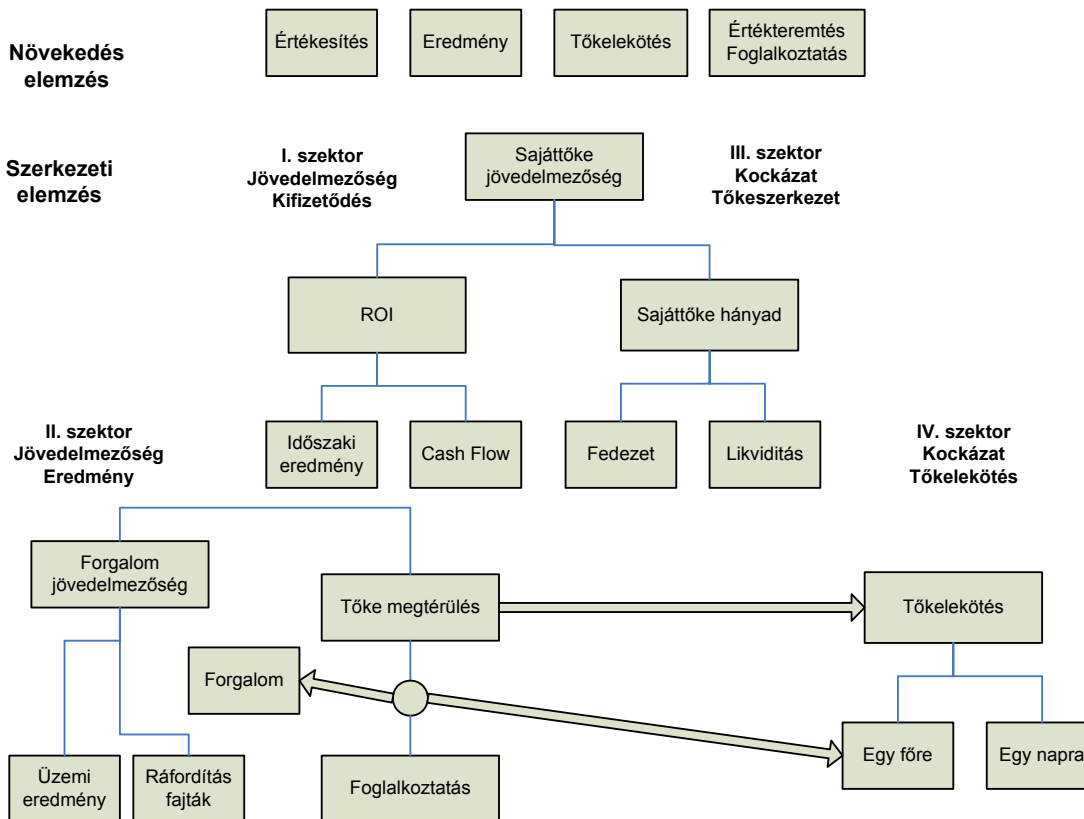
A mutatószámok eredményes használatának egyik feltétele, hogy rendszerben legyenek. Az első, azóta széles körben elterjedt és továbbfejlesztett ilyen rendszert a DuPont vegyi vállalatnál vezették be sok évtizeddel ezelőtt. Ez alapvetően pénzügyi célokat szolgál. A legfontosabb részét a 4.3. ábra mutatja be.

Ezt a rendszert fejlesztette tovább a ZVEI (Zentralverband der Elektroingeneure) német mérnökegyesület a szerkezeti elemzés irányába történt kiterjesztéssel. Nyilvánvaló ugyanis, hogy a mutatók által jelzett eltérések okát és kiküszöbölésük módját csak további, szerkezeti elemzésekkel lehet meghatározni.

Ezek csak kiragadott, bár kétségkívül jellegzetes és fontos példák, még számos efféle rendszer létezik, a DuPont-féle rendszernek is számos továbbfejlesztése.



4.3. ábra: A DuPont mutatószám rendszere



4.4. ábra: A ZVEI mutatószám rendszere



## 4.6. A képi megjelenítés (vizualizálás) módszere

### 4.6.1. A vizualizálás szerepe

Amint a 4.3. alpont (A menedzser-barát informatika) bevezetőjében láttuk, a tágabb értelemben vett képi megjelenítésnek (korszerű, és szűkebb értelemben a számítógépekhez kötött kifejezéssel vizualizálásnak) jelentős múltja, hagyománya van, amit a képességek részéről az emberi idegrendszeri folyamatok, a feladatok részéről pedig a társadalmi munkamegosztásban az egyénre háruló követelmények közötti eltérés indokol. Hiszen már jóval a számítógépek feltalálása előtt meghaladta az egyén képességeit az akkor létező összetett rendszerek áttekintése, feladatainak megoldása, ugyanakkor az efféle feladatok elvégzésére kényszerülő emberek hamar rájöttek, hogy alkalmas vizuális segítséggel sokkal eredményesebbé, vagy – sok esetben – egyáltalán képessé tudnak válni a feladatra. A számítógép megjelenése új lendületet adott a kérdéskörnek, és egyúttal ösztönzést a tudományos kutatásra. Ez a kutatás kiterjed az emberi „információfogyasztás”-ra, az emberi információfeldolgozási folyamatokra, a feladatoknak megfelelő módszerek és ábrázolások keresésére, ennek keretében a vizualizáció sok feladatban közös jellemzőinek a feltárására, és az eszközök kutatására, fejlesztésére.

Sok tudományterület alakult ki, amelyek részben egymástól függetlenül fejlődtek, egymástól független szaknyelvet alakítottak ki, és így a művelőik gyakran nehezen értik meg egymást, a kívülálló pedig még kevésbé. Az igény viszont egyre nő az efféle alkalmazások iránt, ezért sok területen vannak már – nem szükségképpen tudományos eredményeken alapuló – használható eszközök (szoftverek), amelyeknek az eredményei viszont, sajnos, általában nehezen vihetők át más területre. A következőkben kísérletet teszünk a témakör áttekintésére, amit a tudományos részletek szabatos tárgyalása helyett inkább néhány példa bemutatásával szeretnénk közelebb vinni az olvasóhoz. Az áttekintés Heckenast Tamásnak „Üzleti információ vizualizálás” címen megjelent tanulmánya alapján készült, amely a Raabe kiadónál „Controllingtrendek” címen megjelent tanulmánykötetben található.

A 4.1. táblázatban összefoglaltuk a tárgykör legfontosabb fogalmait.

#### 4.1. táblázat: Definíciók

Külső kogníció	A külső világ felhasználása kognitív tevékenység elvégzésére.
Vizualizálás	Számítógép alapú, interaktív vizuális reprezentáció a kogníció erősítésére.
Tudományos vizualizálás	Tudományos, fizikai háttérrel rendelkező adatok vizualizálása.
Információ vizualizálás	Absztrakt, fizikai háttérrel nem rendelkező adatok vizualizálása.

Kiegészítésül: a „kognitív” elsődleges jelentése „megismerő”. Ennek megfelelően a „kogníció” megismerést jelent.

#### 4.6.2. A számítógépes vizualizálás rövid története

A vizualizáció szó elsődlegesen valamilyen dolog belső vizuális elképzelését, mentális, tehát tudatunkban megjelenő képének megalkotását jelenti. Az így kialakított képet mentális modellnek tekintve, segítségével további következtetések vonhatók le az adott esetben absztrakt dologra vonatkozóan. A továbbiakban itt nem ebben a jelentésében használjuk a vizualizálás kifejezést, hanem abban az értelmében, melyet valamikor az 1980-as évek közepén kapott a számítógépes grafika bizonyos fajta felhasználásánál. Ilyen értelemben tehát szinonimája a grafikus adat, illetve információ **megjelenítésnek**. Másképpen az a folyamat, mely adatok vizuális képkénti (látványszerű) reprezentációját állítja elő. Legszélesebb értelmében egy számítógépes „eszköz”, mely az adatokból közvetlenül nem látható összefüggések láthatóvá tételét van hivatott segíteni az emberi szem és látókérgi speciális, gyors percepciós és térbeli képességek hatékony kiaknázásával. A továbbiakban a vizualizálás és az adat vagy információ megjelenítés kifejezések felcserélhetők egymással.

A vizualizálás kifejezés első számítógépekhez köthető megjelenése a tudományos számításokhoz, és szuper számítógépes alkalmazásokhoz kapcsolódik. Ezekben az alkalmazásokban igen nagy mennyiségű numerikus számítást végeztek el valamilyen fizikai, tér-időbeli dinamikus rendszerrel kapcsolatban (légműködés, orvosi képképzés, mérnöki szerkezetek méretezése, földtudományok, csillagászat). Ekkor jelenik meg a tudományos vizualizálás (scientific visualization) fogalma, mint egy olyan technikáé, mely elősegíti a tudományos felfedezést és megértést, valamint

kommunikációs és oktatási közegként működik. Ekkor fogalmazódnak meg a tudományos vizualizálással kapcsolatos kritériumok és jelennek meg önálló kutatási témaként. Az egyik leglényegesebb kitétel, hogy ezek a vizualizációs megoldások túlmutatnak a nyomtatásban kommunikált grafikus reprezentációkon abban, hogy felhasználóik interakcióba léphetnek az adatokkal, egy felhasználói felületen keresztül módosítva a megjelenítés paramétereit. Ez az interakciós követelmény az, mely sokáig gátolta, hogy a kifejlesztett megoldások széles körben elterjedjenek, ugyanis az interaktív megjelenítés megkívánta számítási kapacitás sokáig még csak nagy teljesítményű speciális munkaállomásokon és szuper számítógépes környezetben állt rendelkezésre. Ez természetesen korlátozta a felhasználó tudományterületeket is. Az önálló területként való megjelenés egybe esik a grafikus hardverelemek teljesítményének egy-két nagyságrendes növekedésével. A tudományos vizualizálással kapcsolatban jelenleg is igaz, hogy a legkiforrottabb technológiai eszköztárral és tudományos háttérrel, kutatási módszertannal rendelkezik, mely erősen kapcsolódik a számítógépes grafikához és digitális képfeldolgozáshoz. Napjainkban legalapvetőbb alkalmazási területei a gyógyászat, valamint a járműipari mérnöki tervezés, továbbá – természetesen – a katonai és űrkutatási felhasználás.

Az 1990-es évek elejének személyi számítógépes forradalma magával hozta az uralkodó számítógépes felhasználási paradigma megváltozását. Egyre inkább meghatározóvá válik az interaktív felhasználás. Erősödni kezd az ember-gép kapcsolattal foglalkozó HCI (human computer interaction) kutatási terület. Ezeknek a kutatásoknak egy része az információs munkavégzés új útjait keresi, így a kutatók természetesen kerülnek szembe az információ megjelenítés problémájával. Mivel az általános számítógép használat során általában absztrakt, időnként virtuális fogalmakat, objektumokat és a hozzájuk kapcsolódó adatokat kell megjeleníteni, és ezek manipulációját kell lehetővé tenni a felhasználók számára, a tudományos vizualizálás eredményei csak korlátozottan használhatók. Megjelenik az általános információ vizualizálás (information visualization), mint tudományos terület, mely az absztrakt információk megjelenítésére koncentrálna. Az absztrakt adatok megkövetelnek az alkalmazás architektúrájában egy további lépést, valamilyen vizuális analógia, metafora kialakítását. A továbbiakban az információ vizualizálás kifejezést ebben az értelemben használjuk.

Az információ vizualizálás tipikus alkalmazási területei a pénzügyi adatok feldolgozása, dokumentumkezelő és információ visszakereső rendsze-

rek, statisztikai elemző csomagok. Az ilyen megjelenítő rendszerek tervezésénél a fő nehézséget az adatok tömege és komplex belső struktúrája jelenti. Mivel az egyes alkalmazási területeken egyrészt a megjelenítendő adatok fizikai szerkezete, másrészt szemantikája nagyon eltérő lehet, és ugyanakkor a támogatni kívánt felhasználói feladatok is eltérőek, nehéz az egyes alkalmazási területek sikeres megoldásainak átültetése az egyre újabb és újabb alkalmazási területekre.

Az információ vizualizálás egyik általános funkciója az úgynevezett explorációs<sup>1</sup>, humán felfedezésre épített adatelemzés támogatása. Ebben a helyzetben a humán elemző számára a vizualizáció biztosítja a megfelelő kapcsolódási felületet az adatokhoz, és az elemző az adatokkal folytatott interakcióban alakítja ki mentális modelljeit a valóságról. Természetesen a vizualizáció nem helyettesítheti a klasszikus szigorú kvantitatív számításokra épített elemzést, de segítheti a kvantitatív elemzés fókuszálását, és a kvalitatív irányítást adhat. Azaz segíti az információ kivonatolását, áttekintését, segít a szerkezet, mintázatok és trendek azonosításában. Segítségével azonosíthatók a további érdeklődésre számot tartó területek.

A tudományos vizualizálástól eltérően az információ vizualizálással foglalkozó tudományos közösség általában nem arra fókuszál, hogy egy-egy területre megfelelő megoldásokat szállítson, hanem hogy automatizálja komplex vizualizációk létrehozását és tervezését, illetve általános módszertani elveket dolgozzon ki. Vagyis az adatok szerkezetének, a kommunikációs célnak és a grafikus reprezentáció összhangjának megoldása a cél. Továbbá a HCI közösséggel való nagy átfedés miatt, részben a vizualizációk újfajta felhasználói felületekben való alkalmazására és használhatósági elemzésére koncentrálnak. Összefoglalva tehát, az információ vizualizálás területe egy tipikusan multi-diszciplináris és fiatal tudományág, egyre növekvő alkalmazási területtel.

Az információ vizualizálás leginkább kutatott területe az információ visszakereséshez (information retrieval) köthető. Ennek egyik fontos oka a gigászi méreteket öltő dokumentumkönyvtárak, adatbázisok megjelenése, köztük a World Wide Webbel. Az üzleti vizualizálás területe egyelőre az egyike, a még csak gyermekcipőben járó alkalmazási területeknek. A legfontosabb nehézség, hogy az adatok, ahogyan már korábban említettük, nagyfokúan absztraktok, továbbá sokdimenziósok, valamint a támogatandó felhasználói tevékenység, az üzleti modellezés és döntéshozatal maga is

---

<sup>1</sup> Felfedező, feltáró

bonyolult. Az üzleti információ megjelenítésének igénye természetesen nem új. Régióta használnak táblázatokat, hierarchikus kivonatokat, kör és oszlop diagrammokat. Azonban a modern üzleti vizualizálás üzleti adatok speciálisan megtervezett, többdimenziós interaktív megjelenítését jelenti. A legegyszerűbb döntéstámogató rendszerek egyszerű diagramkészítési képességekkel ellátott lekérdező és jelentéskészítő alkalmazások. A fejlettebb rendszerek absztrakt grafikus elemeket, hierarchikus szerkezetű megjelenítőket, animációkat is magukba foglaló vizuális interfészekkel rendelkeznek. Azaz komplex módon képezik le az adat attribútumokat (tulajdonságokat) a felhasználó által jól felfogható vizuális attribútumokká. Az üzleti adatok egyik fontos jellemzője az időbeliség, mely speciális kezelést kíván a vizualizálás során.

Az üzleti célú többdimenziós vizualizálás esetén az adatok nem biztosítanak keretet a természetes geometriai elrendezéshez. Esetleg, az adatok közti matematikai összefüggések változóinak dimenziószáma szolgálhat elrendező szerkezetül. Ezek az elrendezések az esetek egy részében egymásba ágyazott 3D (három dimenziós) terek megjelenítését kívánják, ami matematikailag korrekt keretet jelentene, de ennek a percepciója (felfogása) nem bizonyult természetesnek és olyan könnyen használhatónak, mint a 2D elrendezéseké. A helyes megközelítést vélhetően a problémamegoldás folyamat- és döntési modelljei nyújtják a hozzájuk illeszkedő vizuális elrendezésekkel.

Az üzleti vizualizáláson belül a legfontosabb előrelépések az adatbányászás (data mining) vizuális támogatása, tőkepiaci elemzések és a vezetői információs rendszerek (EIS) területén érzékelhetők. Ezek az eredmények a statisztika, az adatbázis kezelés, alakfelismerés, gépi tanulás és a mesterséges intelligencia, valamint a számítógépes grafika eredményeinek kombinálásával jöttek létre. Elsődleges cél a korai rendszerek esetén a már említett explorációs (rejtett ismeretet feltáró) adatelemzés lehetővé tétele volt. Az automatikusan végrehajtható algoritmikus ismeretfeltárás mellett egyre hangsúlyosabbá válik a humán irányítás, felügyelet melletti tudáskinyerés.

Az eddig elmondottakból látható, hogy a hatékony üzleti vizualizálás számos megoldandó nehézséggel kell, hogy megküzdjön. Széles szakadék tátong az akadémiai szféra kutatásai és a széles felhasználói kör alkalmazási gyakorlata között. Bár kétségtelenül nő az érdeklődés az üzleti vizualizálás iránt, az új módszerek, vizuális metaforák nehezen kerülnek át a gyakorlati alkalmazás szintjére. A lehetséges felhasználók egy része nincs tisz-

tában az egyre szélesebb körben hozzáférhető, megnövekedett grafikus számítási kapacitás nyújtotta lehetőségekkel. A helyes utat az érintett közösség olyan speciális multi-diszciplináris képzése jelentheti, ahol találkozik az információ vizualizálás az üzleti modellezés és elemzés problémáival. Ezáltal szélesíthető az alkalmazók tábora és kialakulhat egy magas szintű felhasználói kultúra.

#### 4.2. táblázat: A vizualizáláshoz kapcsolódó tudományterületek

Adatbázis-kezelés
Tudáskinyerés és feltárás (Adatbányászás)
Számítógépes grafika
Képfeldolgozás
Grafikai tervezés, design
Látáskutatás, kognitív tudomány, kognitív pszichológia
Ember-számítógép kapcsolat (HCI) kutatása
Tudományos vizualizálás
Információ-vizualizálás

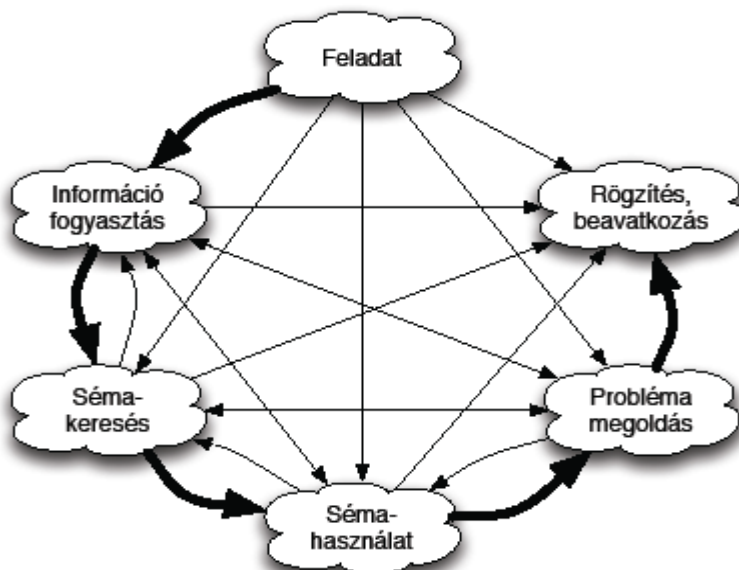
#### 4.6.3. A tudományos modell keretei

A tudáskinyerési folyamatban egy adott személy valamilyen céllal információt gyűjt, szerez és tesz értelmezhetővé egy séma vagy keret kialakításával és tesz azután valamilyen kommunikáció vagy akció formájában elérhetővé. Ez a kimenet lehet egy leírás, szóbeli tájékoztatás vagy egy döntés. Az egész folyamatot nagy mennyiségű inhomogén információ, rossz problémastrukturáltság, de viszonylag jól meghatározott cél jellemzi. A lényeg, hogy a cél elérése megköveteli az információk belátását, felfogását. A tudáskinyerési tevékenység általában része egy összetettebb tudás alapú munkának, például tervezésnek.

A tudáskinyerési folyamat tipikus elemei a következők:

- információfogyasztás,
- sémakeresés,
- sémahasználat,
- problémamegoldás vagy kiértékelés,
- tudás kimenet rögzítése.

Az információfogyasztást egyelőre tekintjük a lényeges információk beszerzésének. A séma megtalálása egy alkalmas reprezentáció, mentális elrendezés vagy modell kialakítását jelenti. Az adatok, nyers információk kódolt formában jelennek meg a sémában. Ez a kódolás levágja az adatok egy részét. Ha ez a veszteség fontos adatokat is tartalmaz, egy jobb sémát kell keresni. Az absztrakciónak ez a formája minden információ feldolgozási folyamatra jellemző, a kezelhetetlen mennyiségű információt a mentális számítási műveletek számára hatékonyan kell redukálni (csökkenteni). Az információ szelektív figyelmen kívül hagyása jellemző az összes biológia információ feldolgozó rendszerre. Az érzékszervek egyszerűsítik, és megfelelően elrendezik a bemeneteket, hogy a magasabb szintű feldolgozó egységeket aggregált (összesített) formában lássák el információval. A magasabb szintű feldolgozásra jellemző a mintázatok detektálása, és az információ mennyiségének további jelentős csökkentése, sokkal absztraktabb és univerzálisabb formákban való további feldolgozása. A séma használata tulajdonképpen ez a lépés, az összegyűjtött információkra. A problémamegoldás végül olyan részfeladatok elvégzését jelenti, melyek a meghatározott minták alapján, az adatok rendezését, csoportosítását, összehasonlítását eredményezik.



4.5. ábra: Tudáskinyerési folyamat

Az információ vizualizálás egyszerűen elősegíti a detektálható és absztrakt mintázatok létrehozását. Az információ vizualizálási megoldások az adatokat leképezik vizuális formákra, melyek az elsődleges sémát fogják alkotni. Az interaktív vizsgálat további összefüggéseket tárhat fel a séma és adatelemek között. A vizuális elemek manipulálásával létrehozott kompaktabb leírások újabb sémáknak tekinthetők. Ezt a vizuális kompaktságot fejezi ki a jól ismert szólás: egy kép többet mond ezer szónál. Azok a megoldások tekinthetők tehát jónak, melyek lehetővé teszik valamely feladat elvégzéséhez megnövelt mennyiségű információ reprezentációját.

Az információ vizualizálás valóban képes hatékonyan támogatni, erősíteni a kognitív folyamatokat. Egyrészt megnöveli a problémamegoldási folyamat számára rendelkezésre álló erőforrásokat:

- Kihasználhatóvá teszi, a humán vizuális figyelmi rendszer biztosította nagy sáv szélességű hierarchikus interakciós csatornát.
- A különböző vizuális attribútumok párhuzamos feldolgozást tesznek lehetővé.
- Megfelelő átkódolással a szimbolikus<sup>2</sup> kognitív feldolgozás áthelyezhető a percepciós<sup>3</sup> (érzékelési) feldolgozás szintjére, ezzel tehermentesíthető a szimbolikus feldolgozó rendszer.
- A külső vizuális reprezentációk kiterjesztik a szűkös munkamemóriát.
- A vizualizálás felhasználható arra, hogy masszív információ mennyiségeket tároljon könnyen hozzáférhető formában.
- Másrészt csökkenti a keresések számát és a kereséssel töltött időt:
  - A vizualizálások az együtt felhasználandó információ csoportosításával, és az ezáltal elért lokalizált feldolgozással csökkentik a kereséseket.
  - A kis térben megjelenített nagy mennyiségű adat nagy adatsűrűséget eredményezhet.
  - Az egy adott objektum köré csoportosított adatok feleslegessé teszik szimbolikus címkék használatát, térbeli indexeléssel teszik lehetővé a címzést.

---

<sup>2</sup> Szimbolikus = jelekkel, szavakkal, valamilyen nyelv szabályai szerint végzett.

<sup>3</sup> Percepció = befogadás, felfogás. Percepció = befogadás, érzékelés során, nem szóbeli, fogalmi gondolkodás útján végzett.



- Továbbá a vizualizálások hatékonyabbá tehetik mintázatok felismerését.
  - A vizualizálás generálta információk felismerése könnyebb, mint a felhasználó általi visszaidézés, emlékezés.
  - A vizualizációk egyszerűsítik és elrendezik az információkat, ezáltal biztosítják a magasabb szinten aggregált és absztrahált információkat a szelektív elhagyással.
- A vizuális elrendezések rendkívül hatékonyak a strukturális összefüggések, minták, sémák megjelenítésében. Képesek kiemelni az értékek közt lévő rejtett kapcsolatokat, trendeket.

A vizualizálás segíti a percepció következtetést is:

- A vizuális reprezentációk lehetővé teszik számos esetben a percepció következtetést, mely sokkal könnyebb az emberek számára.
- A vizualizálással kivitelezhető komplex grafikus számítások.
- A megfelelően tervezett vizuális interfészek a figyelmi folyamatok sajátosságainak kihasználásával biztosítják nagyméretű eseményterek hatékony monitorozását.
- Mivel a vizualizálás interaktív, manipulálható médiumot szolgáltat a statikus diagramokkal szemben, lehetséges a paraméterterek vizsgálata és a mélyben meghúzódó összefüggések felismerése.

#### 4.6.4. Az információ költség-struktúrája

Az információ költség-szerkezete hatásának a leírására a legmegfelelőbb modell a Xerox PARC kutató intézetében kidolgozott, úgynevezett információfogyasztási (information foraging) elmélet. Ez az elmélet az információ felhasználást, fogyasztást állítja párhuzamba az élelemmegszerzés, gyűjtés, fogyasztás biológiai evolúciós-ökológiai modelljével. A modell meglehetősen jól magyarázza a humán információ feldolgozás sok sajátosságát, valamint alkalmasnak bizonyult tervezési keretként való felhasználásra újfajta felhasználó felületek és információ vizualizálási megoldások tervezésében is.

Az adatszerkezetekből, műveletekből és erőforráskényszerekből álló reprezentációk az elvégzendő feladatoktól függő költségstruktúrákat definiálnak. A költségstruktúrák egyik legfontosabb jellemzője az információk fellelési helyeinek száma, valamint a különböző helyekre jellemző információ hozzáférési költség. Mindenféle információs tevékenység esetén jellemzően úgy van kialakítva a munkatér, hogy a költség szerkezetet optimalizálják: a közvetlenül vagy gyakran használt viszonylag kis tömegű

információ könnyen hozzáférhető, a közelben helyezkedik el, míg a nagy tömegű, kevésbé gyakran használt információ költségesebben hozzáférhető, távolabbi adattárakban található meg.

Az információfogyasztás elmélete a biológiai optimális élelem-megszerzési elméleten alapul. A modell a következő alapfeltevéseken nyugszik: léteznek döntési lehetőségek arra vonatkozóan, hogy egy adott területen mennyi időt érdemes a gyűjtögetéssel eltölteni, vannak módszerek a választási lehetőségek értékelésére, és vannak a feladatból, az interfészekből, és a populáció képességeiből fakadó korlátok, melyek meghatározzák a döntések értékét. Általánosan tehát mindenfajta tevékenység elemezhető az erőforrások értékei és a megszerzhető értékek alapján. Konceptcionálisan, az optimális információfogyasztó arra törekszik, hogy maximalizálja az egységnyi költségre jutó értékes információ mennyiségét az adott információs feladat keretei között.

A fogyasztási modelleken belül a folt-modellek (patch model) akkor kapnak szerepet, ha a környezetben az élelem eloszlása nem egyenletes, „foltos”. A fogyasztónak ideje egy részét foltok-közötti kereséssel kell töltenie, hogy gazdag élelem-foltokat találjon. Ha egy ilyet talál, akkor kezdődhet a folton-belüli fogyasztás, illetve meg kell hozni azt a döntést, hogy meddig érdemes a folton belül maradni, és mikor kell elhagyni egy másik foltot keresve. Ez az analógia használható az információfogyasztás esetén is: általában meg kell keresni azokat a releváns (odatartozó, mértékadó) információkban gazdag területeket, melyeken érdemes az információkat feldolgozni.

Ennek az egésznek a szemléltetésére tekintsünk egy példát: tegyük fel, hogy valamilyen információkat kell beszereznünk a Webről. Ekkor a foltok-közötti mozgást végezhetjük egy keresőgép segítségével, míg a folton-belüli feldolgozásra a site-on belül böngészést, a közvetlen linkek követését vesszük igénybe. A foltok-közötti illetve a folton-belüli tevékenységek javítására, nyereségnövelésre a nyomkövetés és a gazdagítás jöhet szóba. Az információs nyom (info scent) követése azt jelenti, hogy a közeli, érzékelhető információ alapján eldöntjük, érdemes-e egy távoli foltot felkeresni. Ez egyfajta heurisztikus<sup>4</sup> keresésnek tekinthető. Ha a nyom megfelelően „erős”, akkor a fogyasztó képes korrekt döntéseket hozni. Ha nincs nyom, akkor a fogyasztó gyakorlatilag véletlen bolyongást végez az információtérben. A gazdagítás vagy dúsítás akkor lehetséges, ha a környezet befolyásolható, formálható. Vagyis csökkenthető az odatartozó információk megtalá-

---

<sup>4</sup> „Heurisztikus”: intuitív, tudományosan nem vagy csak részben megalapozott.

lási, feldolgozási költsége. Ez általában megfelelő szűrő mechanizmusok alkalmazását jelenti. A csökkenő haszonnal jellemezhető folton-belüli fogyasztás idejére Charnov határhaszon elmélete szolgál előrejelzéssel. Eszerint addig kell a folton belül maradni, míg a folton-belüli fogyasztási nyereség sebessége egyenlő az egész környezetre jellemző átlag fogyasztási nyereség sebességével. Az elmélet valamint a folton-belüli és a foltok-közötti javítási lehetőségek jól szemléltethetők költség-nyereség diagramon.

Az információ vizualizálás magas szintű célja tehát a fogyasztási nyereség növelése, vagyis annak a biztosítása, hogy a felhasználó ugyanannyi információhoz kevesebb idő alatt jusson hozzá, vagy ugyanannyi idő alatt több információhoz jusson. Az elmélet azt is lehetővé teszi, hogy a különféle vizualizációs megoldások hatékonysága összehasonlítható legyen.

#### 4.6.5. Vizuális problémamegoldás

Tisztázandó kérdés, hogy az információfogyasztás során milyen kognitív erőforrások milyen korlátokat jelentenek. Ugyanis e korlátok eltolása, vagy legalábbis figyelembevétele eredményezhet csak hatékony vizualizálást. A kognitív feldolgozás gazdaságosságának legfontosabb korlátozott erőforrása a figyelem. Ha ezt a feladat szempontjából érdektelen vizuális zaj köti le, vagy az információ megjelenítésének üteme nem illeszkedik az emberi mintázat-feldolgozás sebességéhez, a rendszer rosszul fog működni. A hatékony vizualizálás egyrészt támogatja a vizuális mintázat felismerést, vagy vizuális lekérdezést, másrészt memória kiterjesztésként szolgál. Durva közelítéssel a vizuális feldolgozás során háromféle memória érintett: az ikonikus, a rövidtávú vagy munkamemória és a hosszú távú memória. Az ikonikus memória egy nagyon rövid élettartamú reprezentációja a retinán kialakuló képeknek. A vizuális munkamemóriára jellemző, hogy elkülönült a verbális (szóbeli gondolkodási) munkamemóriától, korlátos kapacitással rendelkezik mindössze három-öt objektum befogadására, az objektumok pozíciója egocentrikusan tárolódik, figyelmi folyamatok befolyásolják a működését. Ezeknek a sajátosságoknak a figyelembevételével kialakítható egy vizuális problémamegoldási eljárás. Ennek eleme az adott vizualizáció.

A problémamegoldás első lépése a probléma kontextusának, követelményeinek, előzetes lépéseinek meghatározása. Ez a lépés nagyrészt nem vizuális. A következő lépés a vizuális lekérdezés megkonstruálása. Ez a probléma olyan átformálását jelenti, mely megengedi a vizuális megoldást. Ezután egy mintázatkeresési ciklus kezdődik, mely a feladat megoldásához fontos elemi vizuális minták megkeresésére irányul. Ennek a kivitelezése

az ún. triádikus modell keretei között történik. Eszerint az anatómiai, élet-tani és lélektani megfigyeléseken nyugvó modell szerint, a feldolgozásnak három lényeges szintjét kell elkülöníteni. A legalsó szintet a korai (figyelem nélküli) feldolgozási szint jelenti, mely gyorsan, párhuzamos folyamatokkal képes részletes átmeneti proto-objektumok létrehozására. A második rendszer egy korlátozott kapacitású figyelmi rendszer, mely ezekből a struktúrákból stabil objektumreprezentációkat hoz létre. A harmadik rendszer egy szintén korlátozott kapacitású rendszer, mely a kontextus biztosításával irányítja a figyelmet a látvány megfelelő objektumaira a látvány absztrakt jelentése és térbeli elrendezése alapján.

Ez a megközelítés több követelménnyel is szolgál az információt megjelenítő rendszerekre vonatkozóan. Az adatokat olyan formában kell megjeleníteni, hogy az információt hordozó minták érzékelése, felfogása könnyű legyen, azaz a vizuális feldolgozó rendszer mintázatkereső képességét kell kihasználni. Tulajdonképpen az információfogyasztási elmélet szerinti gazdagítást kell biztosítani. A felhasználói felület kognitív befolyását minimalizálni kell, a gondolatoknak a probléma és nem az interfész körül kell forogni. Továbbá az interfésznek az alacsony költségű, gyors információkezesésre optimalizálnak kell lenni. Ez utóbbi követelmény a vizualizációkban alkalmazott navigáció költségeinek csökkentését jelenti, ami visszanyarodva a fogyasztási elmélethez, az információs nyom növelését jelenti.

Összefoglalva, a legjobb vizualizációk dinamikus objektumok, melyek képesek a felhasználó különböző nézetek és részletesség iránti igényeit kielégíteni. A vizualizálás egy kétirányú interfész, amely természetesen aszimmetrikus, sokkal nagyobb sávszélességgel a géptől az ember felé, mint fordítva. E miatt az aszimmetria miatt, a kogníciót támogató rendszereknek fél automatikusan kell működniük, esetenként felhasználói irányítással. Az interaktív vizualizálás összeköti és kombinálja az inputot és az outputot. Minden vizuális objektum megjeleníthet adatokat, és azok manipulálásával fogadhat inputot. A vizualizálás tervezése során két, egymásnak ellentmondó követelményt kell kielégíteni: a legjobb vizuális reprezentációt kell választani, valamint konzisztensnek kell maradni az adott szakterület legtöbbször standardizált fogalmaival.

#### **4.6.6. Felhasználás az üzleti folyamatokban**

Az üzleti vizualizálás sajátossága, hogy nagy mennyiségű, absztrakt, nem geometriai adatot kell megjeleníteni. A bonyolult összefüggések erősen lekötik a döntéshozókat, különösen időkényszer mellett. Egyelőre az üzleti

célú vizualizálás tervezése, fejlesztése mérnöki munka és művészet ötvöze-  
te. Szükség van az „iparos” szakértelmére és intuíciójára egyaránt. Annak  
ellenére, hogy az egyes konkrét vizualizációk szakterület specifikusak, sze-  
rencsére a fejlesztési módszertanok és alkalmazható technikák lehetnek  
ettől függetlenek. Fontos szempont, hogy a tervezés ne technológiai, ha-  
nem humán orientált legyen.

Az üzleti döntések mögötti mentális modellek erősen absztraktak és  
nehezen képezhetőek le geometriai objektumokra. A legtöbb rutinszerűen  
alkalmazott tudományos vizualizálási megoldás csak korlátozottan alkal-  
mazható. Nem egyszerű a főleg információ visszakeresési alkalmazások-  
hoz kifejlesztett információ vizualizálási eljárások átvitele az üzleti alkal-  
mazások területére. Egy adott problémához nem feltétlenül van alkalmaz-  
ható olyan tervezési minta, mely illeszkedik a problémamegoldás mentális  
eljárásaihoz, és mindeközben percepciósan elég erős, hogy erősítse a kog-  
nitív folyamatokat, és megfelelő interakciót is biztosítsa a felhasználó és a  
vizualizáció között.

A vizualizálás három elsődleges helyen használható a modellezés so-  
rán. Tegyük fel, hogy a modellezőnek egy feladat megoldásához fel kell  
fedeznie összefüggéseket nagy mennyiségű absztrakt adat között. Erre  
alapozva létre kell hoznia egy számítási modellt, mely rögzíti a feltárt ösz-  
szefüggéseket. Végül ezt a modellt alkalmaznia kell az adatokra validálási  
célokból vagy más adatokra kiterjesztve következtetések levonására. Eb-  
ben a folyamatban jól felhasználható egy explorációs interaktív vizualizáló  
alkalmazás. Segíthet az adatok közti összefüggések felfedezésében, a mo-  
dell létrehozásában, tulajdonságainak megismerésében és a modellel elvég-  
zett számítások eredményeinek tanulmányozásában. A vizualizálás segíti  
továbbá több modellező kooperatív munkáját, a modellezők és a végfel-  
használók kommunikációját.

A legnehezebb kérdés üzleti célú vizualizálás tervezésekor a megfelelő  
vizuális objektumok, vizuális metafora megtalálása. Ez részben abból fa-  
kad, hogy nem nyilvánvaló, mi a természetes reprezentáció üzleti adatok  
számára. Léteznek ugyan szokásos reprezentációk, de a korrekt metafora  
megtalálásához nem sok segítséget találunk. Az üzleti döntéshozók számá-  
ra jól ismertek az üzleti diagramok. Napjainkban az oszlop diagramok,  
grafikonok, kör diagramok „természetesnek” tekinthetők. Valójában ezek  
nem természetes, hanem tanult formák. Ezek a formák persze speciális  
elrendezésekkel kiterjeszthetők több dimenziós adatok kezelésére.

Az üzleti alkalmazások első és leggyakoribb területe a **kontrolling** támogatása. Ennek a legmagasabb stratégiai szinten kialakított rendszere az **eredménykártyák** (scorecards) használata. Ezek fő célja az üzleti stratégia és a végrehajtás összhangba hozása. Itt magas szintű, több mutatóból kialakított KPI-k (key performance indicator) megjelenítése a cél, általában a felső vezetés számára. A legmodernebb üzleti intelligencia szoftvercsomagok vezetői moduljai tartalmazznak eredménykártya megoldásokat.

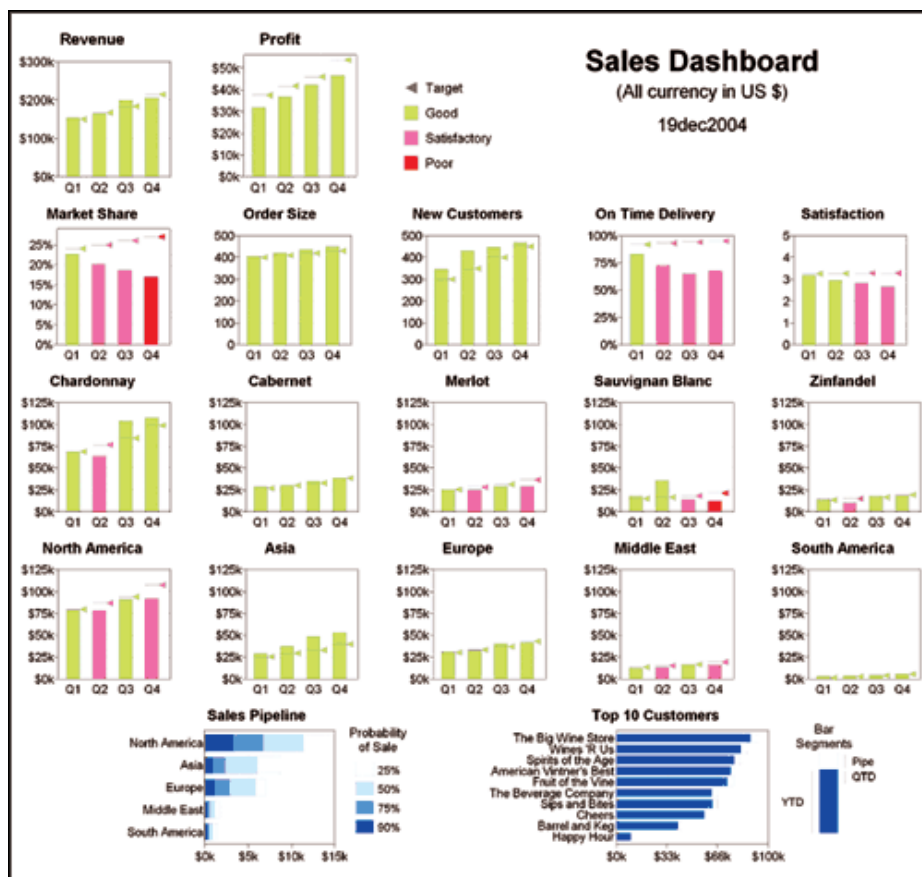
Az **üzleti műszerfalak** (dashboard) egy szinttel lejjebb találhatóak, ezek inkább közvetlen működési célokhoz kapcsolódnak. Fő céljuk a beavatkozásokhoz szükséges üzleti információk nyújtása intuitív és átlátható formában.

A műszerfal rendszerek top-down tervezése alapvetően eltérő az üzleti intelligencia megoldások bottom-up tervezésétől. Először meg kell határozni az üzleti célokat, az ezekhez tartozó mutatókat és KPI-eket, majd ezek előállításához szükséges adatforrásokat. Ebbe a folyamatba be kell vonni a műszerfalak felhasználóit, hiszen csak ők tudják meghatározni a döntési folyamat számára hasznos, specifikus üzleti adatok körét.

A következő, 4.6. ábra bemutat egy példát.

Újabban egyre nő az igény a **tőzsdei információk** megjelenítésére. Mindnyájan tudjuk, hogy meglehetősen kaotikus adathalmazról van szó, amelyben azonban két irányból erős az áttekintés igénye. Erős egyrészt a felhasználók részéről, akik szeretnének megalapozott befektetési döntéseket hozni. Ezek közé tartoznak a spekulánsok is, akik a tőzsde működésének szándékos megzavarásával akarnak haszonhoz jutni. Főként ez utóbbi csoport miatt, de a kapitalizmus rendeltetésszerű működése végett is, az üzemeltetők és a felügyeleti hatóságok szeretnék biztonsággal elkülöníteni a rendes, kaotikus mozgást és a gerjesztett, tehát rendellenes mozgást.

Korábban már hivatkoztunk az iparban és a közműveknél alkalmazott felügyeleti rendszerekre, amelyekben a humán döntéshozók munkáját segíti egy monitorozó rendszer. Ezek a monitorozó rendszerek az esetek egy részében nemcsak egyszerűen közlik, megjelenítik a felügyelt rendszer állapotát, de (újabbban) intelligens módon értékeli is azt, segítve a humán döntéshozót, operátort. E rendszerek létesítésekor mindig nagy súlyt fektettek a kommunikatív felhasználói felületekre, melyek az esetek többségében információ vizualizálásra épülnek. Ilyen fajta rendszerekre az ipari folyamatirányítás, infrastrukturális felügyeleti rendszerek a legközvetlenebb példák, de találhatunk ilyen rendszert az üzleti élet területén is.



4.6. ábra: Egy műszerfal megjelenítés

A következőkben a NASDAQ felügyeletét ellátó NASD Regulation Inc. által fejlesztett rendszerről lesz szó (4.7. ábra). Maga a NASDAQ nagyon sok elektronikus tranzakciót bonyolít le a különböző földrajzi helyeken lévő tőzsdei felek között. Az említett cég a tőzsde felügyeletét látja el, feladata a meg nem engedett kereskedelmi tevékenység felderítésére és megakadályozására. A rendszer a NASDAQ kereskedelmi tevékenységére, a kereskedők és papírok profiljaira épített adattárházra épül, adatbányászati és információ vizualizálási technikák integrálásával.

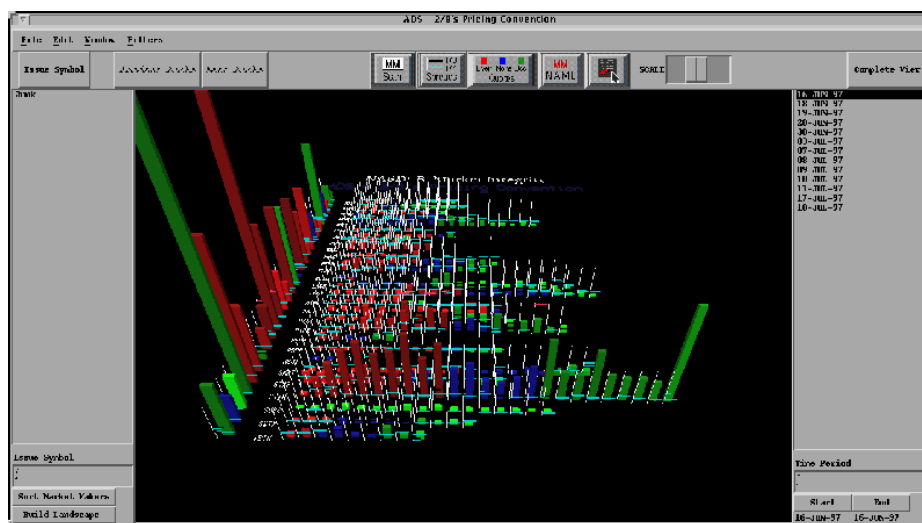
Egy döntési fa és társítási szabály modul generálja a normál kereskedelemre jellemző viselkedési mintákhoz tartozó szabályokat. Ezeket a szabályokat humán elemzők validálják egy grafikus felhasználói felületen keresztül. Ez a felület megjeleníti a szabályok mögötti érintett tranzakciók össze-

foglalását, és biztosítja a tranzakciók részleteihez a lefűrást. Ugyanez a felület szolgál egy esetleges szabálysértéskor az adott szabály, és az azt megsértő tranzakciók elemzésére.

Szabálysértések gyanújakor a rendszer lehetővé teszi az adott időpont körüli piaci állapot tanulmányozását: követhető a részvények áramlása a piaci szereplők között, valamint a piaci szereplők kapcsolatai. A piaci kontextus részét képezik az ajánlati és kötési árfolyamok.

A rendszer 3D vizualizálás segítségével képes megjeleníteni a teljes részvény piacra vonatkozó információkat, például a versenyellenes árképzés detektálására. Ezek a nézetek lehetővé teszik a kötési viselkedésbeli hasonlóságok gyors felderítését, közelebbi részletesebb tanulmányozásukra irányítva a figyelmet.

A rendszer tapasztalatai azt mutatják, hogy az adatbányászás és a vizualizálás kiegészítik egymást. A vizualizálásra szükség van, hogy az elemzők eldönthessék az adatbányászás mely eredményei érdekesek. Az így kialakított minták képezhetik majd egy automatikus detektáló rendszer alapját.



4.7. ábra: Árfolyam térkép

Itt még felhívjuk a figyelmet arra, hogy a **számítógép-hálózatok biztonságának** védelmében is az egyik leghatékonyabb módszer a hálózati folyamatok felügyelete, és a rendellenes jelenségek idejében történő felismerése. Az Athéni Nyári Olimpiai Játékok idején ismeretlen tettesek napokon, talán heteken át lehallgatták több vezető görög kormánytisztviselő-



nek és politikusnak, köztük a külügyminiszternek a mobil telefonját. Több éves nyomozás derítette ki, hogy a büntettet lehetővé tevő több mulasztás egyike az volt, hogy amikor az érintett mobilszolgáltató a számára törvényben előírt lehallgatási rendszert telepítette, a vevő kívánságára nem telepítette hozzá azt a kezelői felületet, amelynek a használatán keresztül a rendellenes működést nagy valószínűséggel az első órákban észrevehették volna.

## 5. Az OLAP technológia

Az adatbázis- és információs technológia az 1960-as évektől kezdve lépésről lépésre, fokozatosan jutott el az alacsony szintű, egyszerű műveletekre épülő fájlkezeléstől a hatékony adatbázisrendszerekhez. Az 1970-es években jelentek meg a relációs adatbázisrendszerek. Ekkor a nagy mennyiségű adat hatékony tárolása, elérése és feldolgozása az **on-line tranzakció-feldolgozás (On-line Transaction Processing, OLTP)** eljárásával történt.

A 90-es évek elején jelentkezett először nagy méretekben igény arra, hogy a hagyományos, aktuális állapot lekérdezésére vonatkozó alkalmazások mellett a múltbeli események elemzése alapján a jövőben várható folyamatok megtervezésére is lehetőség nyíljon. Az újfajta igény kielégítésére jelenik meg az adatok rendezett tárolására az **adatáruházz, adattárház (Data Warehouse)**, az elemzések elvégzésére pedig az **on-line elemző feldolgozás (On-line Analytical Processing, OLAP)**.

### 5.1. Az OLAP Codd-féle kritériumai

E.F.Codd (USA) használta először az on-line analytical processing (OLAP) kifejezést. 1992-ben megjelent cikkében vezeti be az OLAP fogalmát. Az OLAP legfontosabb ismérve, hogy multidimenzionális adatstruktúrát használva lehetővé teszi az adatok gyors és rugalmas lekérdezését, majd ezt követően az adatok analízisét.

Codd cikkében továbbá 12 pontban megfogalmazza az OLAP eszközökkel szembeni követelményeket is. Ez a követelményrendszer azóta általánosan elfogadottá vált, bár egyes kritériumok megváltoztak, illetve újak is kerültek az elvárások közé.

Codd OLAP szabályai:

1. Multidimenzionális nézet: az OLAP termékekben az adatok modellje többdimenziós. Ez lehetővé teszi, hogy a felhasználó igény szerint műveleteket végezzen (adatszeletek kiválasztása, forgatás, lefűrés stb.).
2. Transzparencia (átláthatóság, áttekinthetőség): az eredmény a technikai részletek ismerete nélkül könnyen áttekinthető legyen, a felhasználónak ne kelljen tudnia, hogy az adatok fizikailag vagy logikailag hogyan vannak tárolva.
3. Elérhetőségek (jogosultságok) beállítása.

4. Állandó lekérdezési (jelentéskészítési, reporting) teljesítmény: a felhasználó ne tapasztaljon jelentős teljesítmény csökkenést a dimenziók számának és az adatbázis méretének növelésével.
5. Kliens-szerver architektúra: az adattárházak esetében előforduló hatalmas adattömeget nagy teljesítményű szerverek tárolják.
6. Általános dimenzió fogalom: a dimenzióknak struktúrájukban azonosnak kell lenniük.
7. Dinamikus ritkamátrix-kezelés (sparsity): biztosítani kell a többdimenziós adatmodelleknél előforduló ritkamátrixok (nem teljesen kitöltött) feldolgozásának optimális feltételeit.
8. Többfelhasználós üzemmód támogatása: előfordulhat, hogy pl. több felhasználó ugyanazzal az adatmodellel szeretne dolgozni. Az OLAP eszközöknek biztosítani kell a konkurens (versenyző) elérést, az adatok védelmét, integritását.
9. Korlátozás nélküli dimenzióműveletek: a dimenziók közötti műveletek minden formáját meg lehessen valósítani.
10. Intuitív adatkezelés: a felhasználók számára az adatok manipulálása közvetlenül (segédeszköz nélkül) a megjelenítési felületen megvalósítható legyen.
11. Rugalmas jelentés (report) készítés: támogatott legyen az adatok legkülönbözőbb megjelenítése.
12. Korlátlan dimenziószám és aggregációs szintszám: az OLAP eszköz legyen képes tetszőleges számú dimenzió és hierarchia szint kezelésére. Természetesen a valóságban ez nem képzelhető el. Codd szerint legalább 15 szint kezelése az elvárható, de inkább 20 dimenzió kezelése a kívánatos. Tapasztalatok szerint konkrét esetben ritkán van szükség 8-10 dimenzió kezelésénél többre.

Újabban a **FASMI (Fast Analysis of Shared Multidimensionál Information)** betűszóval szokták megadni az elvárások összességét.

Ez egyrészt azt jelenti, hogy a felhasználó gyorsaságot (fast) vár el a lekérdezéseknél- ez egyszerűbb eseteknél 5 másodpercet jelent, de bonyolultabb esetekben se kelljen 20 másodpercnél többet várni.

Másrészt az OLAP rendszerek képesek legyenek tetszőleges, a felhasználó számára fontos üzleti vagy statisztikai számítást, elemzést (analysis) elvégezni. Ezekhez általában semmilyen, vagy legfeljebb minimális programozásra legyen szüksége a felhasználónak.

Harmadrészt megosztott (shared) legyen az OLAP eszköz, ami azt jelenti, hogy teljesítse az adatok bizalmas kezeléséhez szükséges követelményeket, és képes legyen kezelni az adatok több felhasználó általi egyidejű elérését.

Utoljára pedig az OLAP eszköz kulcsfontosságú tulajdonsága, a többdimenziós (multidimensional) adatstruktúra használata szerepel a rövidítésben.

## 5.2. Az OLTP és OLAP rendszerek összehasonlítása

Az adatbázisok hagyományos alkalmazási területei közé sorolhatók a raktári nyilvántartások, termelési adatok tárolása, számlák kezelése, könyvtárak nyilvántartási és kölcsönzési rendszere stb. Ezek közös jellemzője, hogy az adatbázis a vizsgált rendszerek adminisztrálására szolgál, az adatbázisban a rendszer leírása, az aktuális állapota található. Az ilyen rendszerekhez tartozó alkalmazásokra az jellemző, hogy a lekérdező funkció mellett képesek a valós folyamatoknak megfelelően megváltoztatni, aktualizálni az adatbázis tartalmát.

Pl. egy raktári nyilvántartó rendszerben az adatbázis a raktár aktuális feltöltöttségét mutatja meg, de megadja azt is, hogy melyik termékből mennyi van raktáron, hol helyezkedik el stb.. A raktárnál előfordulnak ki- és beszállítások. Mindkettő megváltoztatja a raktár állapotát. Ezért a rendszerhez tartozó alkalmazásnak olyan elemeket is kell tartalmaznia, amelyek képesek kezelni (leképezni) ezeket a változásokat, hogy az adatbázis mindig az aktuális adatokat tartalmazza.

Az OLTP rendszerek feladata a fent leírtak megvalósítása, azaz az online tranzakciók és lekérdezések elvégzése.

Az egyre élesedő piaci verseny szükségessé tette a naprakész információkra épülő, gyors, de ugyanakkor megalapozott döntéseket. A döntéshozáshoz elengedhetetlen a múltbeli események elemzése, a várható trendek felismerése. Ehhez olyan adatbázisra van szükség, amely nem csak az aktuális állapotot tartalmazza, hanem a múltbeli adatokat is megőrzi és több, különböző forrásból származó nagy mennyiségű adatot tartalmaz. Ezen igényeknek megfelelő adatbázis az adattárház. Az OLAP rendszerek lényege nagymennyiségű adaton végzett elemzés orientált adatkezelés.

Az OLTP és az OLAP rendszerek közötti legfontosabb különbségek a következők:

1. Felhasználók szerint: Az OLTP **ügyfélorientált**, általában egy vállalat ügyintézői és informatikusai használják tranzakciók és lekérdezések végrehajtására. Az OLAP viszont **piacorientált**, döntéshozók és az őket segítőik használják az adatok elemzésére.
2. Adatok jellege szerint: Az OLTP rendszerek a vizsgált terület **aktuális állapotát leíró adatokat** tartalmaznak, amelyek túl részletesek ahhoz, hogy döntéshozatalnál könnyen alkalmazni lehessen őket. Az OLAP rendszerek ezzel szemben nagymennyiségű, **időrendben archivált adatot** kezelnek. A döntéshozatal megkönnyítése érdekében az OLAP rendszerek lehetőséget nyújtanak az adatok különböző szinteken való összegzésére és tárolására.
3. Adatbázis szerkezete szerint: Az OLTP rendszerekre a hagyományos **relációs adatmodell** (egyed-kapcsolat (EK) adatmodell) és az **alkalmazás-orientált adatbázis-használat** a jellemző. Az OLAP rendszerek jellemzően **csillag vagy hópehely adatmodellt** használnak, és **témaorientált adatbázis-használat** jellemzi őket.
4. Adatok származása szerint: Az OLTP rendszerek általában **egyetlen adatforrásra** építenek, a szervezet aktuális, egy homogén modell szerint tárolt adatait használják. Ennek számos előnye van a fejlesztési és karbantartási tevékenységek vonatkozásában. Az OLAP rendszerek különböző **adatforrások adatait integrálják**, használva a korábbi és más szervezetektől származó adatokat is.
5. Rendszerhez való hozzáférés szerint: Az OLTP rendszerek esetében a működés szokványos módja a több, konkurrensen futó rövid ideig tartó, **elemi tranzakció**. Ezek a tranzakciók rendszerint adatmódosítási műveleteket tartalmaznak. Az OLAP rendszerekhez való hozzáférést a csak olvasó (read-only) műveletek jellemzik. Ezek a **lekérdezések** általában **komplexek**, egy döntés meghozatalánál ugyanis az a jó, ha a problémát sok oldalról, többféle megközelítésből is megvizsgáljuk.

Az OLTP és OLAP rendszerek összehasonlítására más jellemzőket is szokás használni. Ilyen pl. az adatbázis mérete, műveletek gyakorisága stb. Az 5.1. ábra tartalmazza az OLTP-OLAP összehasonlításra leggyakrabban használt jellemzőket.

Tulajdonság	OLTP	OLAP
funkció	operatív feldolgozás	döntéstámogatás
felhasználó	ügyintéző	elemző
használat	rendszeresen ismétlődő	alkalmi
adatok	naprakész	történeti
összegzés szintje	nagyon részletes	összesített
elérés	írás/olvasás	főleg olvasás
elért rekord szám	tízes nagyságrend	millió nagyságrend
felhasználók száma	ezres nagyságrend	száz nagyságrend
adatbázis méret	100 MB-tól	100 GB-tól
teljesítmény mutató	tranzakciós idő	lekérdezési idő

5.1. ábra: Az OLTP és OLAP rendszerek összehasonlítása

### 5.3. A többdimenziós (multidimenziós) adatmodell

Az OLAP terminológiák és műveletek jelentősen eltérnek a relációs algebrai műveletektől és terminológiáktól. Ezért célszerű olyan adatmodellt használni, amelyek jobban megfelel az OLAP igényeinek.

Miért is van szükség a másféle adatmodellre? Tekintsük azt az esetet, amikor egy kereskedelmi tevékenységgel foglalkozó vállalat vezetőjének kell döntést hoznia pl. arról, hogy cége melyik régióban, milyen termékeket reklámozzon, hogy eladásait növelni tudják, hol van lehetőség új egységek nyitására ahol a forgalom már kinőtte a jelenlegi kapacitásokat stb.

A megfelelő információ ilyen helyzetben elengedhetetlen. A vezető számára azonban nem minden információ megfelelő. Az eddigi összes eladás listája túl részletes, sok olyan adatot is tartalmaz, amelyek a fontos részleteket eltakarják. A vezető ilyen esetben összesítő információkat igényel.

A jó döntéshez nem csak pusztán tények ismerete, hanem a különböző tényezők közötti kapcsolatok ismerete is szükséges. A fent említett döntéshez nem elegendő az előző évekre vonatkozó éves szintű eladási adatok ismerete egy-egy termékre vonatkozóan, jól használható lenne pl. két termék, vagy több régió egymáshoz viszonyított forgalma. Szükség lenne tehát e mennyiségek kapcsolatának ismeretére, a kapcsolat alakulásának ismeretére.

Olyan adatmodell tehát a megfelelő, amely úgy tárolja az adatokat, hogy minél könnyebben le lehessen kérdezni a különböző mennyiségek közötti kapcsolatokat. A multidimenziós adatmodell megfelel ezeknek az elvárásoknak.

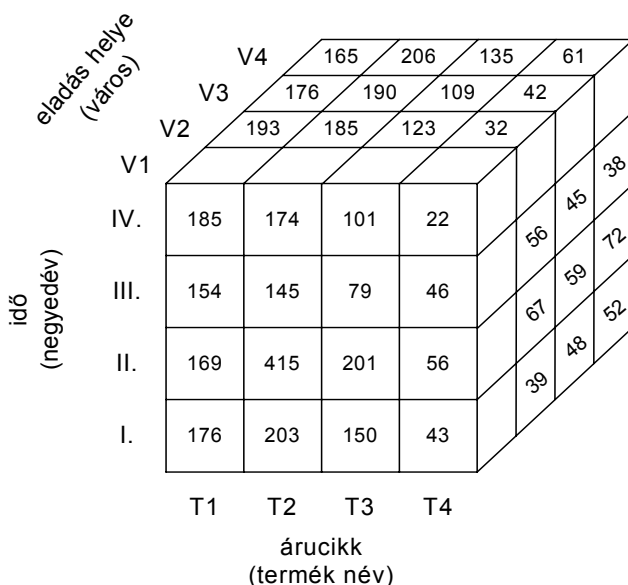
Az OLAP eszközök multidimenziós adatmodellt vesznek alapul, amely az adatot adatkockának tekinti

### 5.3.1. A multidimenziós adatkocka

Az adatkockát az éleihez rendelt **dimenziók (dimension)** és a belsejében, a cellákban tárolva a **tények számértékei (measure)** határozzák meg.

Azt mondhatjuk, hogy a dimenzió mindig egy, az elemzés szempontjából lényeges nézőpont, szempont, a tények számértékei pedig azok a mennyiségek valamilyen mértékegységben megadva, ami szerint elemezni akarjuk a dimenziók közötti összefüggéseket.

Az OLAP alapvetően témaorientált, ezért a többdimenziós adatmodell jellemzően valamilyen meghatározott, az elemzés szempontjából fontos téma köré szerveződik. Egy kereskedelmi vállalat esetében ez természetesen lehet az értékesítés.



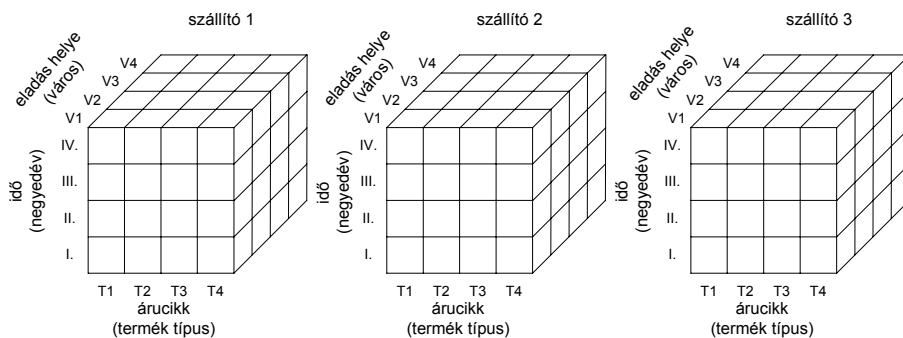
5.2. ábra: Háromdimenziós adatkocka

Ezért az értékesítés adatait tartalmazó adattárház esetén dimenzió (elemzési szempont) lehet az idő, árucikk, szállító, eladás helye például. Ezekkel a dimenziókkal lehetővé válik például annak vizsgálata, hogy az egyes árucikkek esetén hogyan alakult a negyedéves forgalom, mely városokban volt adott negyedévben kimagasló az értékesítés egy adott árucikkből stb.

Példánkban a tényértékek lehetnek az értékesítés adatai valamilyen pénzegységben megadva, vagy az értékesített darabszám árucikkenként, de lehet az eladási egységár is.

A kocka szó hallatán elsődlegesen a három dimenziós geometriai alakzat jut az eszünkbe. Ha a felsorolt dimenziók közül egyenlőre csak az idő, árucikk és eladás helye dimenziókat, tényadatként pedig az eladott darabszámot tekintjük, akkor a modellünk valóban egy 3 dimenziós adatkockával adható meg, amelyet az 5.2. ábra mutat.

Ha további elemzésnél az egyes szállítóktól történő beszerzésre vonatkozó adatok ismerete is szükségessé válik, akkor a negyedik, a szállító dimenziót is figyelembe kell venni. A négy dimenziót ábrázolni a hagyományos módon nem lehet, de a négy dimenziós kockát felfoghatjuk úgy, mint három dimenziós kockák „sorozatát”. Ezt mutatja az 5.3. ábra.



5.3. ábra: Adatok szemléltetése a négydimenziós kockában

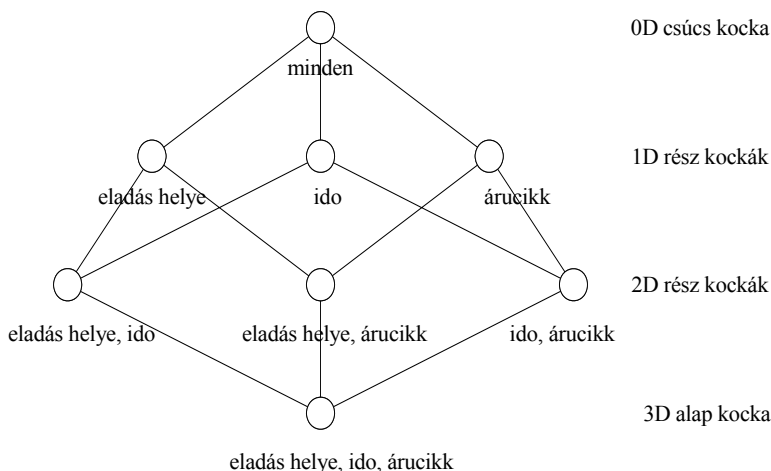
Ebből is látható, hogy az adatkocka tulajdonképpen egy elnevezés a többdimenziós adattárolásra, az adatok valódi fizikai tárolása eltérhet a tárolás logikai reprezentációjától.

A fent bemutatott kockák az adatokat különböző összegzési szinten tartalmazzák. Az ilyen adatkockákat **rész-kockáknak** nevezik. Ismerve a dimenziók halmazát, megadható a **rész-kockák hálój**a, ahol minden rész-kocka az adatok másfajta összegzését (azaz a dimenziók egy más részhal-



maza szerinti összefoglalását) mutatja. Ha a dimenziók halmazának az eladás helye(város), idő(negyedév) és árucikk(termék típus) dimenziókat tekintjük, akkor az 5.4. ábra mutatja a rész-kockák hálóját. Ez a háló az, amelyet tulajdonképpen ehhez a három dimenzióhoz tartozó adatkockaként kezelünk.

Azt a rész-kockát, ahol a legalacsonyabb szintű az adatok összegzése, **alapkockának** nevezik. Az összegzés legmagasabb szintje pedig a **csúcs-kockában** található, ez egyetlen adatot jelent. A közbűső szinteken található rész-kockák adják az OLAP rendszerek alapkövetelményeként megfogalmazott aggregált adatokat. Hogy mikor hány és milyen szintű rész-kockát érdemes kiszámítani és fizikailag is tárolni, arra a későbbiekben még visszatérünk.

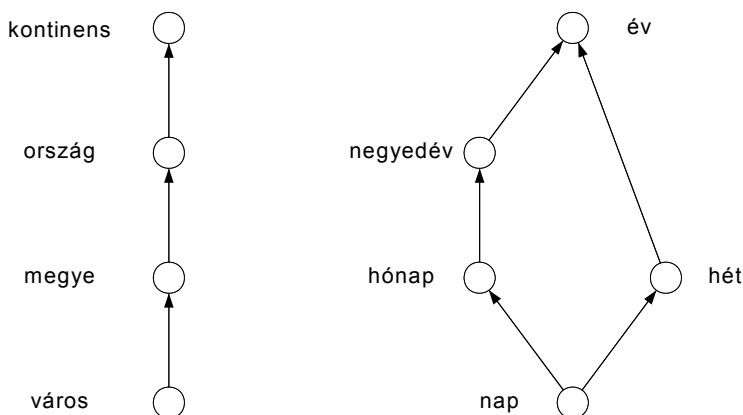


5.4. ábra: Egy háromdimenziós adatkockát alkotó rész-kocka háló.

### 5.3.2. Dimenzió hierarchiák

Egy dimenzió alacsonyabb és magasabb szinteket tartalmazhat. Például az eladás helye dimenzió esetén a városok Magyarországon területileg egy-egy megyéhez sorolhatók, a megyék egy országhoz rendelhetők. Multinacionális cég esetén folytathatnánk a hozzárendelést az országoknak kontinensekhez való hozzárendelésével. Így egy leképezés sorozatot adtunk meg, az alacsonyabb szintű fogalmaktól (város) a magasabb szintű, általánosabb fogalmakig (kontinens) haladva. Ezt a leképezés sorozatot nevezzük hierarchiának.

Vannak dimenziók, ahol magától értetődő a hierarchia. Ilyen pl. az idő dimenzió. Ha az idő dimenzió attribútumai a nap, hét, hónap, negyedév, év, akkor a nap<hét<hónap<negyedév<év teljes rendezéssel is kialakítható egy hierarchia, de részleges rendezést is lehet alkalmazni, amikor egy hálóstruktúrájú hierarchia rendszer adható meg. Ezeket mutatja az 5.5. ábra.



5.5. ábra: Dimenzió hierarchiák

Egy adott dimenzióhoz több hierarchia is megadható a különböző felhasználói igények szerint. A hierarchiákat létrehozhatják szakértők, statisztikai elemzések alapján automatikusan generálhatók, de manuálisan maguk a felhasználók is megadhatják.

### 5.3.3. Műveletek a többdimenziós adatkockán

Az eddigiek alapján láttuk, hogy a többdimenziós adatmodellben az adatok dimenziók mentén szerveződnek, az egyes dimenziókban létrehozott hierarchiának megfelelően. Az OLAP műveletek ezeket a hierarchiákat használják, lehetővé téve azt, hogy a felhasználó különböző szinteken vizsgálhassa az adatokat.

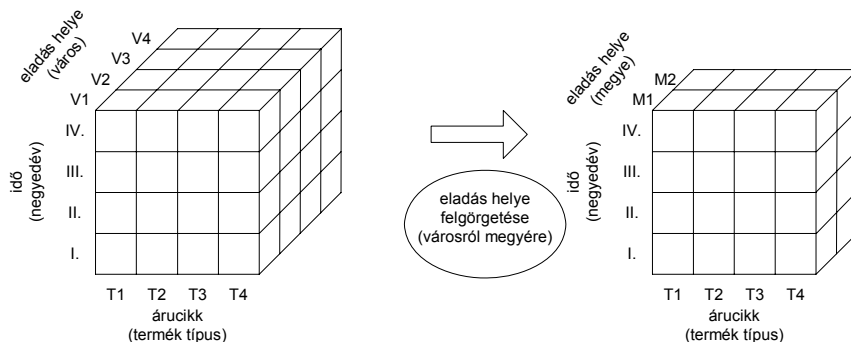
A továbbiakban ismertetésre kerülő néhány tipikus OLAP művelet bemutatására az 5.2. ábra által bemutatott adatkockát használjuk.

## Felgörgetés (roll up)

A felgörgetés egy összesítő művelet, amelynél az összesítés egy adott dimenzió hierarchiája mentén történő felfelé haladással, vagy dimenzió csökkentéssel hajtódik végre.

Az 5.6. ábra azt az esetet mutatja, ha a kiinduló adatkockából az eladás helye dimenziónál a város, megye, ország, kontinens hierarchiát feltételezve a felgörgetés a város szintről a megye szintjére történő emelkedés esetén összesíti az adatokat. Az új kockában városok helyett megyék szerint vannak csoportosítva az adatok.

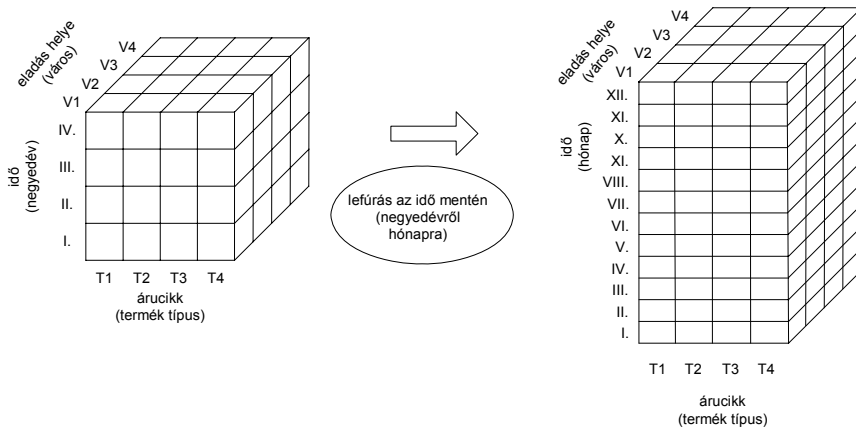
A dimenziócsökkentéssel végrehajtott felgörgetésnél dimenzió(ka)t hagyunk el. Például ha az eladott mennyiséget tartalmazó adatkockát tekintjük szállítók, eladás helye, idő, árucikk dimenziókkal, akkor a felgörgetést a szállítók dimenzióra végrehajtva – összesítve az összes szállítóra az adatokat – az 5.2. ábra szerinti adatkocka adódik.



5.6. ábra: Felgörgetés

## Lefűrés (drill down)

A lefűrés a felgörgetés művelet ellenkezője, itt a kevésbé részletes adatoktól haladunk a részletesebb adatok felé. Ez kétféleképpen valósítható meg: vagy egy adott dimenzió hierarchiája mentén egy szinttel lejjebb lépünk, vagy új dimenzió bevezetésével. Az 5.7. ábra azt a fűrészt mutatja, amikor az idő dimenzió mentén lépünk egy szinttel lejjebb, az adatokat negyedéves bontás helyett havi bontásban tartalmazza az adatkocka.

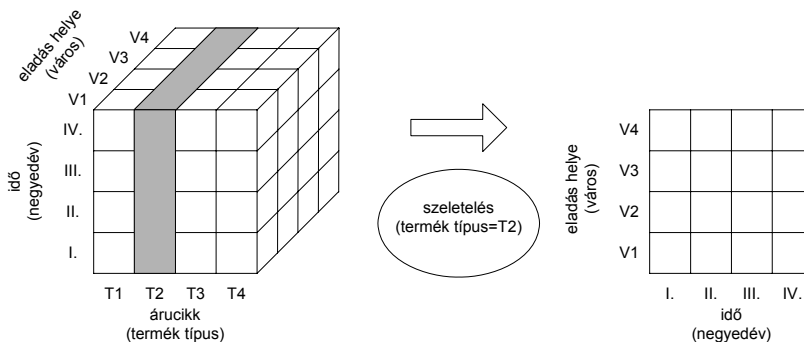


5.7. ábra: Lefűrés

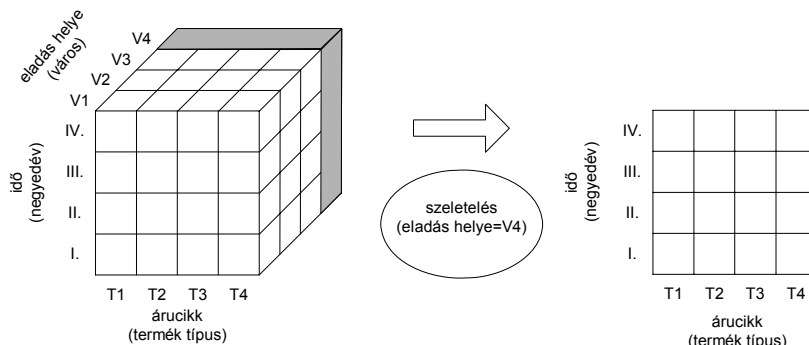
### Szeletelés és kockázás (slice and dice)

Mindkét művelet egy-egy kiválasztási műveletet hajt végre.

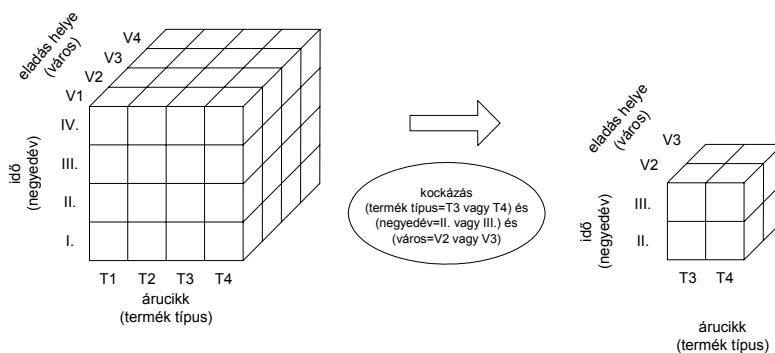
A szeletelés egy adott kocka adott dimenziója mentén történő kiválasztást jelent. A T2 termékért felelős termékmenedzser szempontjából az 5.2. ábra szerinti kocka T2 termék típusra vonatkozó adatai fontosak, míg a V4 város eladásaiért felelős területi menedzsert az 5.2. ábra adatkockájának V4 városhoz tartozó szelete érdekli. Ezeket a szeleteket mutatja az 5.8. ábra és 5.9. ábra.



5.8. ábra: Szeletelés a termékmenedzser szempontjából



5.9. ábra: Szeletelés a területi menedzser szempontjából



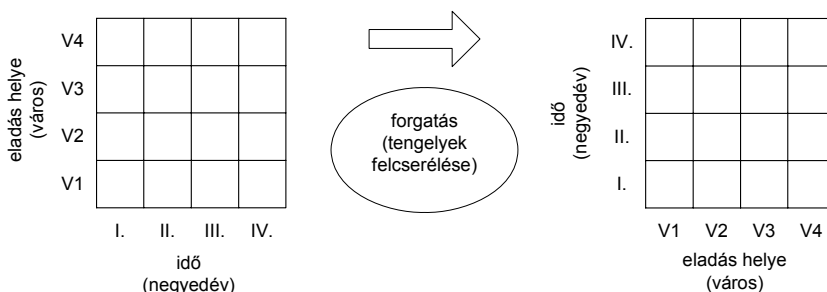
5.10. ábra: Kockázás

### Forgatás (pivot)

A forgatás művelet lehetővé teszi, hogy a jelentésekben az adatok a lekérdező számára megfelelő, szemléletes formában jelenjenek meg. Ez a művelet nem az adatok halmazát, hanem csak az adatok képi megjelenését változtatja meg. Pl. az 5.8. ábra szerinti két dimenziós szelet esetében a pivotálás az adattengelyek felcserélését jelenti. Ezt mutatja az 5.11. ábra. A pivotálás speciális esetének tekinthető, amikor egy háromdimenziós kockát kétdimenziós síkokba transzformálunk.

Egyes OLAP rendszerek más fúrési műveleteket is támogatnak. Lehetővé teszik az átfúrást, a keresztlufúrást. Pl. átfúrás segítségével valósítható meg egy olyan lekérdezés, amely egyszerre több tényadatot tartalmaz.

OLAP alapfunkcióként szokás megadni egyéb műveleteket is, mint pl. a különféle mozgó átlagok számítását, gazdasági növekedési mutatók számítását, és egyéb statisztikai számítások elvégzését.



5.11. ábra Forgatás

#### 5.3.4. Hatékony adatkezelés

Az adattárházak rendkívül nagymennyiségű adathalmazából az elemzők nagyon rövid idő (másodpercek) alatt kívánnak lekérdezéseikre választ kapni. Ehhez elengedhetetlen, hogy a többdimenziós adattárolás hatékony legyen. Ez a hatékonyság két részből tevődik össze: egyrészt az aggregált adatok kezeléséből, másrészt az adatkocka kitöltöttségének, ritkaságának kezeléséből.

Az 5.3.1 fejezetben említésre került, hogy az összesített, aggregált adatokat a részkockák tartalmazzák. Ha a részkockákat vagy egy részüket előre kiszámítjuk, akkor ez nagyon gyors válaszidőt eredményez, és segítségével felesleges (ismétlődő) számítások is elkerülhetővé válnak.

Fontos kérdés annak eldöntése, hogy milyen mértékű legyen a részkockák előszámítása. Egy  $n$  dimenziós kockához  $2^n$  részkocka tartozik, beleszámítva az alapkockát is. Ahogy az 5.4. ábra mutatja, 3 dimenzió esetén 8 részkocka adható meg. Ha a dimenziókhöz hierarchia is tartozik, akkor ez az érték sokkal nagyobb lesz,

$$\prod_{i=1}^n (L_i + 1)$$

alakban adható meg, ahol  $L_i$  az  $i$ -edik dimenzióhoz tartozó szintek száma. Ha például 4 a szintek száma egy 10 dimenziós kocka minden dimenziója esetén, a részkockák száma  $5^{10} \sim 10^7$ .

Ebből látható, hogy az összes részkocka előszámítása és eltárolása nem reális. Ebben az esetben ugyanis a nagyon nagy tárhely igény mellett az is előfordulhat, hogy a kiszámított részkockák közül számosat egyáltalán nem is fognak a lekérdezések során használni.

Sok és nagyméretű részkocka esetén ésszerűbb megoldás a részleges megvalósítás. Ez azt jelenti, hogy a részkockák csak egy részhalmazát hozzák létre, vagy egy-egy hierarchiában nem minden szintet számítanak ki. Ez kompromisszumot jelent: kevesebb a tárigény, de hosszabb lehet a válaszidő, hiszen ekkor a lekérdezés futásidejében történik a kihagyott szintekhez tartozó érték számítása. A kiszámításra kerülő részkockák kiválasztásánál célszerű figyelembe venni az előforduló lekérdezéseket, azok gyakoriságát.

Egy adatkocka celláinak száma az egyes dimenziók hierarchia szintjein található elemek szorzata. Ha 100 termék eladását vesszük 4 megyében, megyénként 5-5 városban (24 dimenzió elem) éves, negyedéves és havi adatok (17 dimenzió elem) esetében, akkor ez  $100 \cdot 24 \cdot 17 = 40800$  adatcellát jelent. Mivel valószínű nem minden termék kerül eladásra minden időszakban, ezért sok adatcella üres lesz. Ha ezeket az üres cellákat nem tároljuk le, akkor jelentős tárterület takarítható meg. (Ez a probléma különösen fontos sok dimenzió és nagy kiterjedésű dimenzió esetén.) A ritkaság kezelésére egyes multidimenziós adatbáziskezelők tartalmazznak ún. ritka mátrix algoritmust, amely a kocka szerkezetéből megpróbálja a nem használt részeket kiszűrni. Pl. a kockának csak azok a „lapjai” kerülnek tárolásra, amelynek van olyan cellája, amelyhez érték is tartozik.

### 5.3.5. Sémák a multidimenziós adatmodellhez

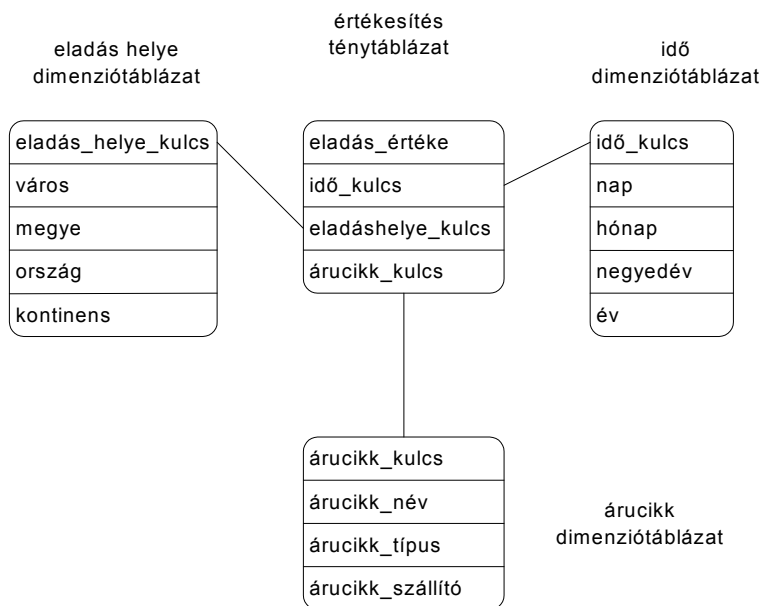
Az OLAP technológia alkalmazásához az adattárházakban az adatok fizikai tárolása történhet relációs vagy többdimenziós formában. Az adattárolás mikéntje alapján szokás beszélni relációs OLAP, vagyis ROLAP, illetve multidimenziós OLAP, vagyis MOLAP rendszerekről. Harmadik lehetőség a HOLAP, azaz hibrid OLAP rendszer alkalmazása. Ekkor a kettő kombinációja valósul meg: nagymennyiségű részletes adat relációs adatbázisban tárolódik, miközben az összesítéseket külön MOLAP tárban tárolják.

A ROLAP rendszerekben az adatok tárolása hagyományos relációs adatbáziskezelővel történik. Gyakori alkalmazása rugalmasságának, a relációs adatbáziskezelők viszonylagos olcsóságának és megbízhatóságának, és a relációs tárolási technika elterjedtségének köszönhető.

A leginkább használt adattárház sémák a csillag séma és a hópehely séma.

A **csillagséma (star schema)** a leggyakrabban alkalmazott modellező séma. A sémagráf egy csillaghoz hasonlít: középen helyezkedik el egy nagy méretű, redundáns adatokat nem tartalmazó központi táblázat (a ténytáblázat), amely meghatározza a többdimenziós modell dimenzióit és ezt veszik körül a kisebb – dimenzióként egy – táblázatok (a dimenziótáblázatok).

A következő ábra csillag sémára mutat példát egy kereskedelmi cég értékesítési adataihoz, ha az értékesítést 3 dimenzió – idő, hely, árucikk – mentén tekintjük



**5.12. ábra:** Csillagséma az értékesítési adatokhoz

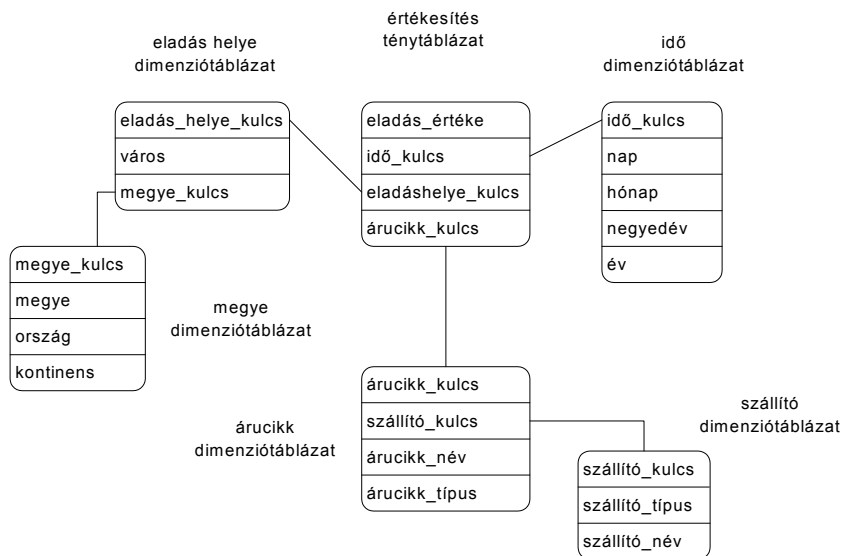
A séma központjában az értékesítéshez tartozó ténytáblázat áll. Ez a táblázat kétféle adatot tartalmaz: mind a három dimenzióhoz kulcsokat és az eladás értéke tényadatot. (Célszerű a ténytáblázat méretének minimalizálása miatt generált kulcsokat használni.) A dimenzió táblák a tényadatok



jellemzőit tartalmazzák dimenzióként összegyűjtve, melynek minden elemét egy kulcs azonosítja. Ezek a táblák általában redundáns adatokat is tartalmaznak. Pl. ha az eladás helye dimenzió táblázat attribútumai a város, megye, ország, akkor két ugyanabban a megyében és országban fekvő városhoz tartozó bejegyzés redundanciát eredményez.

A csillag séma előnye, hogy egyszerű adatmodell, használata kevés tábla olvasását igényli, kevés join adatbázis műveletet igényel és a modell metaadatai (adatokat leíró adat) egyszerűek. Hátránya viszont az aggregált adatok nehézkes képzése, a redundáns adattárolás és nagy dimenzió táblák esetén a hierarchia kezelés lassú lekérdezést eredményezhet.

**Konzolidált csillagsémának** nevezik azt a speciális csillagsémát, amikor a központi ténytáblában aggregált adatokat is tárolnak.



5.13. ábra: Hópehelyséma az értékesítési adatokhoz

A **hópehelyséma (snowflake schema)** a csillagséma egy olyan változata, amikor a ténytáblához közvetlenül csatlakozó dimenzió tábla további altáblákra van osztva. Ezzel lehetővé válik a dimenzió táblák normalizálása, ami a szükséges tárhely csökkenését és a redundancia megszüntetését, vagy mindenképpen a csökkentését jelenti. Ezen előnyök mellett hátránya viszont ennek a sémának, hogy a táblák számának növekedése miatt növekedhet az adatlekérdezések bonyolultsága. Egy lekérdezés végrehajtásához

több összekapcsolás műveletre (join) lesz szükség. Ezért ez a séma kevésbé használatos az adattárház-tervezésben. Ez a séma is a sémagráf alakjáról kapta a nevét.

### 5.3.6. Elemzési folyamat az OLAP segítségével

Egy amerikai, ruházati cikket forgalmazó kereskedelmi vállalat 13 városban rendelkezik üzlettel. Ismertek a cég 1999-2001 években történt eladásaira vonatkozó adatok.

A cég vezetői értékeli és elemzik a 3 éves periódus eladásait. Céljuk a három legnagyobb eladási bevételt elért termék-vonalnál vizsgálni az eladási folyamatokat, különös tekintettel az esetleges eladási visszaesésekre, ha szükséges, javaslatot tenni bizonyos termékek értékesítésének megszüntetésére.

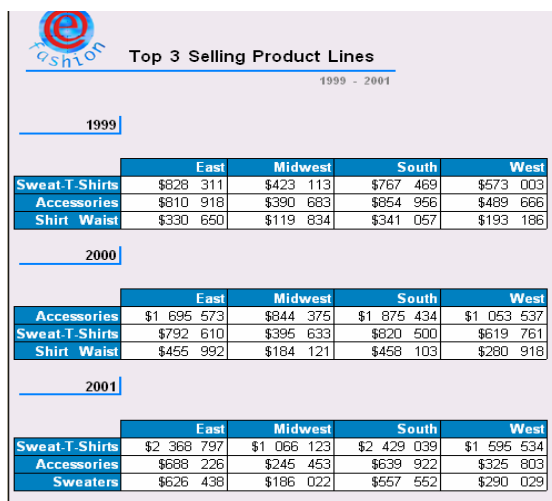
Az elemzésre használt részkoeca jellemzői:

A részkoeca dimenziói:

idő ( hét (week) < hónap (month) < negyedév (quarter) < év (year) ),  
 eladás helye ( város(city) < régió (regio) < állam (state)),  
 termék ( SKU fajta (SKU desc) ) < kategória (category) < termék vonal ( lines) ).

A tényérték: eladás bevétele \$-ban (sales revenue).

A következőkben az elemzés menetét mutatjuk be.

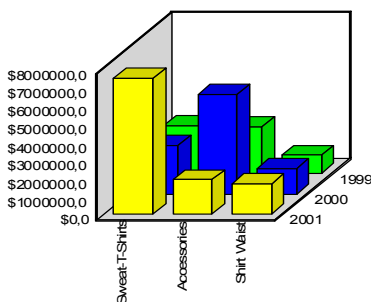


Top 3 Selling Product Lines					
1999					
	East	Midwest	South	West	
Sweat T-Shirts	\$828 311	\$423 113	\$767 469	\$573 003	
Accessories	\$810 918	\$390 683	\$854 956	\$489 666	
Shirt Waist	\$330 650	\$119 834	\$341 057	\$193 186	
2000					
	East	Midwest	South	West	
Accessories	\$1 695 573	\$844 375	\$1 875 434	\$1 053 537	
Sweat T-Shirts	\$792 610	\$395 633	\$820 500	\$619 761	
Shirt Waist	\$455 992	\$184 121	\$458 103	\$280 918	
2001					
	East	Midwest	South	West	
Sweat T-Shirts	\$2 368 797	\$1 066 123	\$2 429 039	\$1 595 534	
Accessories	\$688 226	\$245 453	\$639 922	\$325 803	
Sweaters	\$626 438	\$186 022	\$557 552	\$290 029	

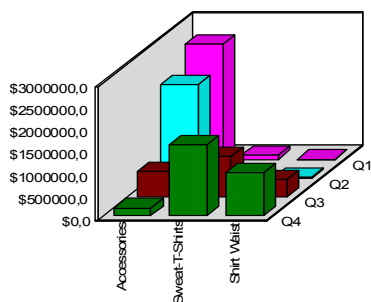
Ebben a táblázatban évenkénti bontásban régióként láthatók az eladási értékek. Mivel időbeli változás vizsgálata a cél, alakítsuk a táblázatot át ennek megfelelően, a táblázat csak évenkénti bontásban mutassa az eladási értékeket (pivot műveletet és egy felgörgetés)!

Top 3 Selling Product Lines			
1999 - 2001			
	1999	2000	2001
Sweat-T-Shirts	\$2 591 895	\$2 628 504	\$7 459 493
Accessories	\$2 546 222	\$5 468 919	\$1 899 405
Shirt Waist	\$984 727	\$1 379 134	\$1 654 358

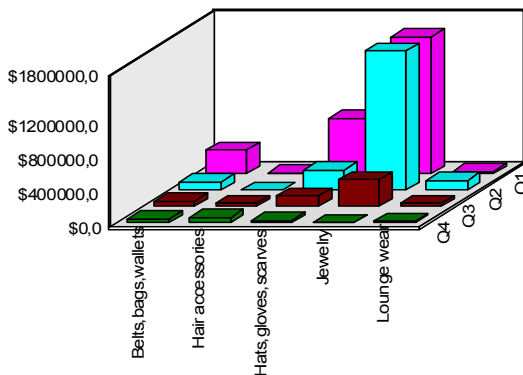
Most jobban látszanak a vizsgált termékeknél a tendenciák, de még szemléletesebben látszanak egy diagramban.



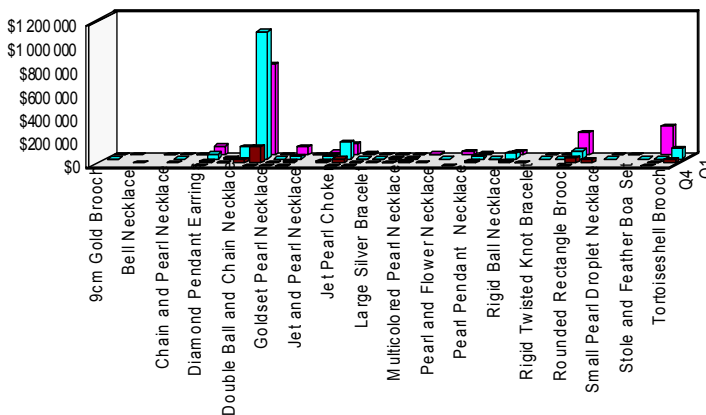
A legszembetűnőbb az, hogy a kellékek (Accessories) esetén 2000 kimagasló értékesítése után 2001-ben nagyon lecsökkent az eladás. Nézzük meg, mi történt 2000-ben. Fúrjunk le az idő dimenzióan egy szinttel lejjebb, lássuk az adatokat negyedéves bontásban.



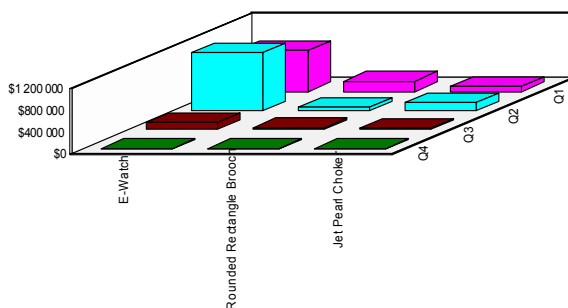
Itt az látszik, hogy a kellékeknél negyedévről negyedévre folyamatosan csökken az értékesítés. Adjuk meg, hogy a kellékeken belül „mi a felelős” a csökkenésért. Folytassuk ezért vizsgálatainkat áttérve egy másik, a termék dimenzióra.



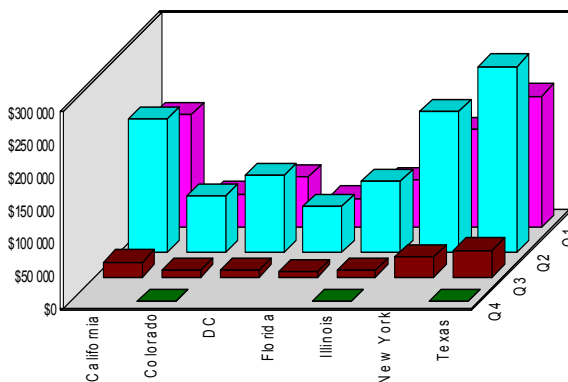
Az újabb fúrás után jól látható a „felelős”, ez pedig az ékszer (Jewelry) kategória. Ezen a dimenzión folytatva a vizsgáldást, egy szinttel lejjebb fúrva további részleteket deríthetünk ki.



Az ékszeren belül nagyon sok típus van, ezek közül látható, hogy kevés az, ahol tényleg lényeges csökkenés történt, az eladásokban. Szűrjük le azokat a típusokat, amelyek az eladási érték felső 10%-át adják.



Összegezve vizsgálódásunkat azt állapíthatjuk meg, hogy egy óratípus (E-watch) az, amelynek vásárlása drasztikusan visszaesett. De vajon területileg hogyan alakult ennek a terméknek a forgalma?



Dimenziót váltva, az eladás helyét nézve ennek a terméknek minden államban igen jelentősen visszaesett a forgalmazása. Valószínű javaslat a termék további forgalmazásának megszüntetése.

A bemutatott elemzés egy kiugró eltérés (csökkenés) okának megkeresését prezentálta. (Az elemzés a Business Object 5.1 Quick Tour modulja alapján készült.) Ugyanígy nyomozhatunk kulcsfontosságú teljesítménymutatók után, figyelhetjük más kritikus tényezők alakulását is.

## 6. Adatbányászás

Napjainkban a cégeknek rohamosan növekvő adatmennyiséget kell kezelniük. Ez azt eredményezi, hogy egyre nehezebbé válik az adatokból a kívánt információ kinyerése. Az érdeklődés ezért fordult az olyan új technikák felé, amelyek lehetővé tették, hogy megismerhetővé, kinyerhetővé váljon a nagy adathalmazokban rejlő tudás.

Az adatbányászás (**Data Mining**) az 1990-es években jelent meg az üzleti köztudatban. **Az adatbányászás olyan adatelemzési folyamat, amely nagy adatbázisokból rejtett, előzőleg nem ismert, hasznos információkat tár fel.** A rejtett összefüggések felismerése és kinyerése mellett igen fontos, hogy adatbányászati módszerekkel szakértői döntések automatizálhatók, ami lehetővé teszi emberi erőforrás megtakarítását és kérdésekre adott válaszok reakcióidejének lerövidítését.

A statisztika is az adatelemzés módszere, de lényeges különbség van a statisztika és az adatbányászás között. A legalapvetőbb eltérés, hogy a statisztika elsősorban mintákból következtet az eredeti sokaság tulajdonságaira, az adatbányászati módszerek pedig a rendelkezésre álló adatbázist elemzik. Másik különbség, hogy a statisztika az adatbázison kívüli információra támaszkodva állítja fel a modelljét, hipotézisét és teszteli azt. Az adatbányászás közvetlenül az adatbázisból képes előzőleg nem ismert összefüggések kinyerésére.

### 6.1. A tudásfeltárás folyamata és az adatbányászás

1995-ben Montreálban tartották az első tudásfeltárásról szóló konferenciát. Itt tették azt a javaslatot, hogy az **adatbázisokban végzett tudásfeltárás (Knowledge Discovery in Databases, KDD)** azt a teljes folyamatot jelentse, amelynek során az adatokból kinyerjük az információt. Azt is leszögezték ezen a konferencián, hogy az adatbányászás a tudásfeltárás folyamatának az a lépése, amelyben az adatokban lévő összefüggések felfedezése történik. Természetesen, mindez nagymennyiségű, emberi úton közvetlenül nem áttekinthető adatok halmazára vonatkozik.

A KDD folyamatát a következő lépésekkel szokás megadni:

1. **Adatkiválasztás:** az elemzéshez szükséges adatok kiolvasása az adatbázisból.
2. **Adatelőkészítés:** adattisztítás, adatbővítés, adat-transzformáció (kódolás)- a nem megfelelő adatok eltávolítása, esetleg a vizsgálatokhoz hiányzó adatok integrálása, az adatoknak a vizsgálatok számára használható alakra hozása.
3. **Adatbányászás:** olyan eljárás, amely során adatbányászási technikák (kluszterezés (*clustering*), osztályozás stb.) alkalmazásával feltárjuk az ismeretlen, új trendeket, összefüggéseket, ill. mintázatokat (*patterns*).
4. **Jelentéskészítés:** az előző lépés eredményeinek a végfelhasználó számára értelmezhető formában történő megadása.

A fenti folyamat egyes lépései általában különböző feladatköröket ellátó emberek együttes munkája. Pl. az adatelőkészítési lépésben együtt kell dolgozniuk az adatgyűjtőknek az adatbányászokkal, a jelentés készítés pedig az adatbányászoknak a szakértőkkel való közös tevékenységét igényli.

A fenti felsorolás azt sugallhatja, hogy a tudásfeltárás egy lineáris folyamat. Bármely lépésben előfordulhat azonban, hogy az előző lépésben kapott információ nem megfelelő. Ekkor vissza kell térni a megelőző lépés(ek)hez, és módosított adatokkal, paraméterekkel, esetleg más eszközzel kell folytatni a munkát, megismételve egyes lépés(ek)e)t. Általában elmondható, hogy a tudásfeltárás iteratív folyamat.

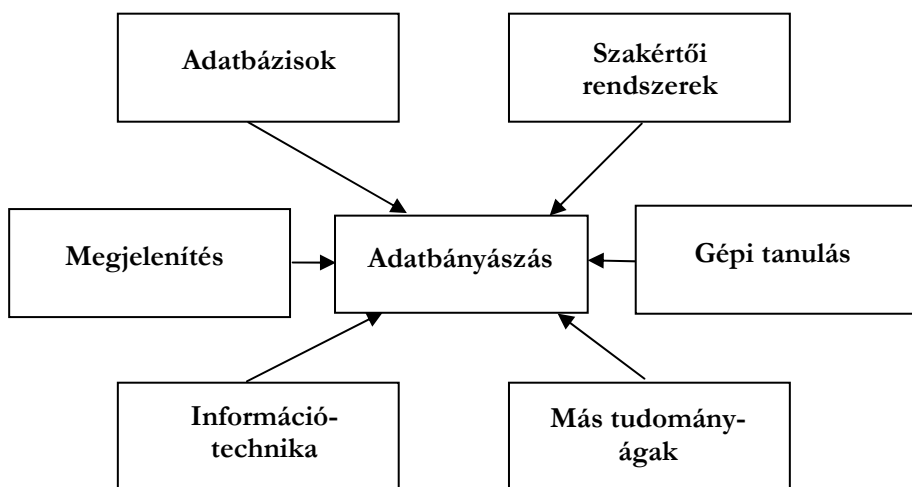
A tudásfeltárás folyamatában az új tudást az adatbányászási lépés adja, ennek ellenére minden lépés egyforma fontossággal bír. Megfelelő adatok nélkül nem lehetséges ugyanis az adatbázisban meglévő ismeretlen összefüggések megismerése. A folyamat időigényét tekintve sem maga a 3. lépés tart a legtovább, a szükséges idő 80%-át az adatelőkészítés és a kapott eredmények értékelhető formába öntése teszi ki.

Az adatbányászás – a fentiek szerint – egy lépés a tudásfeltárás folyamatában. Ennek ellenére az iparban, a felhasználók körében az adatbányászás elnevezést a KDD szinonimájaként is használják. Az adatbányászás terminológia a teljes folyamatra sokkal elterjedtebb, mint a jóval hosszabb „tudásfeltárás adatbázisokban” elnevezés. Ezért szerepelt már a fejezet elején található definícióban is az adatbányászás egy adatelemző folyamatként.

Az adatbányászás interdiszciplináris terület, amely több tudományágot fog egybe, többek között az adatbázisrendszereket, a statisztikát, a gépi

tanulást, a szakértői rendszereket, a megjelenítést és az információtechnológiát. Ezt mutatja a 6.1. ábra.

Az adatbányászás általánosságban bármilyen adatbázisrendszer esetében – relációs adatbázisok, adattárházak, fejlett adatbázisrendszerek, World Wide Web – esetén alkalmazható. A fejlett adatbázisrendszerek közé tartoznak például az objektum orientált és objektum relációs adatbázisok, de ide soroljuk a specifikus adatbázisokat, mint pl. a téradatbázisokat, szöveges adatbázisokat, vagy a multimédia adatbázisokat. Természetesen a különböző adatbázisok esetén különböző adatbányászási feladatok adódnak, amelyeket eltérő adatbányászási technikákkal lehet megoldani.



6.1. ábra: Adatbányászáshoz kapcsolódó tudományágak

## 6.2. Az adatbányászás alkalmazási lehetőségei

Az adatbányászási feladatok – adatelemzési szempontból – általában két osztályba sorolhatók: leíró adatbányászási feladatok és előrejelző (következtetési) adatbányászási feladatok. A leíró adatbányászás az adatbázisban tárolt adatok alap (általános) jellemzőit határozza meg, az előrejelzéses adatbányászás pedig a meglévő adatokból az alapvető összefüggések, mintázatok (adatok egymáshoz viszonyított elhelyezkedése) feltárásával prognosztizál.

Az adatbányászási feladatok megoldásánál a következő technikák a leggyakrabban használtak:



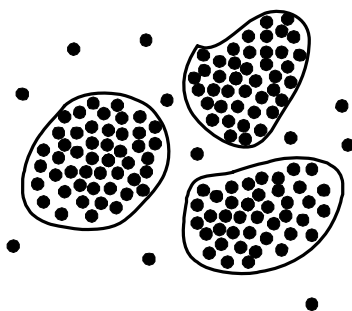
- A **társításelemzés** társítási szabályok (asszociációs szabályok) feltárását jelenti. Ebben az esetben arra vagyunk kíváncsiak, hogy az adatbázis elemei között létezik-e összefüggés. Ha létezik, akkor ez adatbányászási eszközökkel feltárható és a kapcsolat erőssége is jellemezhető. Ez az eljárás széles körben a kereskedelemben, a bevásárlókosár típusú elemzéseknél használható.

Egy adatbányászási elemzés során a következő társítási szabályt tárták fel:

*életkor( XY, 30-40 év) és éves jövedelem (XY, 5-8 millió Ft)  $\Rightarrow$  autót vesz(XY, Audi), (gyakoriság= 5%, bizonyosság=30%),*

ahol XY a vásárló azonosítója. Ez a szabály azt fejezi ki, hogy a vizsgált vásárlóknak (az adatbázisban levő rekordoknak) az 5%-ára érvényes az életkorra és a jövedelemre vonatkozó feltétel és 30 % a valószínűsége, hogy az életkorra és a jövedelemre megfogalmazott feltételeknek eleget tevő vásárlók Audi gépkocsit vásárolnak.

- **Csoportosítás (kluszterezés)** segítségével az adatoknak csoportokba (kluszterekbe) sorolása történik úgy, hogy az egyes csoportokba egymáshoz hasonló elemek kerüljenek, az egyes csoportok viszont jelentősen különbözzenek egymástól. Úgy is szokták mondani, hogy az egy csoporthoz tartozó elemek esetén maximalizáljuk, a különböző csoportokhoz tartozó elemek esetében viszont minimalizáljuk a hasonlóságot. A feltárt csoportok lehetnek egymást kizáróak, de akár egymással átfedők is.



6.2. ábra Kluszterek kialakítása

A kluszterezésre tipikus példa a piackutatás, amikor vásárlási szokások alapján lehetőleg homogén vásárlói csoportokat határoznak meg. Ezek

a csoportok kereskedelmi célcsoportokat reprezentálnak, amelyeket jól lehet használni a marketing tevékenység optimalizálása érdekében: bizonyos reklám anyagot csak a megfelelő célcsoporthoz juttatnak el.

A 6.2. ábra kétdimenziós szemléltetéssel mutatja be az adatok három csoportba (kluszterbe) sorolását.

- Az **osztályozás** az adatbányászás leggyakrabban alkalmazott módszere, mert nem csak a jelenségek leírására, hanem előrejelzésre is használható. Az osztályozás jellemzője, hogy adott az az ismérv, ami alapján az adatbázis elemeit (rekordjait) meg akarjuk különböztetni, osztályokba akarjuk sorolni. Ezt a kiválasztott ismérvet szokás *osztálycímke*-nek nevezni. Osztálycímke csak olyan ismérv lehet, ami véges számú különböző értéket vehet fel. Ez azt jelenti, hogy ismert az is, hány osztály létezik. Az osztályok ismeretében meg lehet adni, prognosztizálni lehet, hogy egy új adatelem (rekord) melyik osztályba tartozik. Az osztályba sorolás osztályozási szabályok alapján történik, amelyeket különféle módszerek segítségével lehet meghatározni. Cél minden esetben olyan szabály felállítása, amelynek segítségével a legpontosabban lehet szeparálni az adatokat a megfelelő osztályba. Az osztályozást gyakran alkalmazzák pénzügyi vizsgálatoknál (ügylek hitelképességének megadása, biztosítási kockázatbecslés stb.), orvosi alkalmazásoknál, szövegbányászatban.

Például egy pénzügyi ügylet hitelképessége szerint szeretné osztályozni jó, közepes és gyenge minőségű osztályokba. Az osztályba sorolás alapján jellemezni lehet az egyes csoportokba tartozó ügyleteket, és ezek alapján a pénzügyi ügylet egy új ügyfél hitelkérelméből azt is el tudja dönteni (előrejelzés), hogy hitel visszafizetés szempontjából jó ügyfél lesz-e.

A csoportosítás és az osztályozás hasonló technikának tűnhet, azonban lényeges eltérés van a két módszer között. Az osztályozásnál az osztálycímke megadásával ismertek az osztályok, annak számossága is, míg a csoportosításnál nem ismertek előre a csoportok, az adatok alapján kell létrehozni az adatokra jellemző csoportokat. Ezek alapján a kluszterezés olyan osztályozásként is felfogható, ahol nem ismert az osztálycímke. Az elmondottak alapján az osztályozás a gépi tanulás területén a felügyelt tanulás (példák alapján történő tanulás) egy formája, míg a csoportképzés felügyelet nélküli tanulási forma.

- A **fejlődésanalízis** az időben változó adatok időben változó viselkedési szabályosságait modellezi, adja meg. A **regresszió-vizsgálat** célja

egy előrejelzésre alkalmas függvény megadása, amelynek segítségével ismert értékekből más numerikus érték(ek)re lehet következtetni. Jó példa erre, amikor értékpapír befektetési döntésekhez az értékpapír-árak alakulásának előrejelzéséhez az értékpapírt kibocsátó társaságok gazdasági fejlődésének jövőbeli szabályszerűségeit tárják fel adatbányászási módszerekkel. Az **idősorok elemzése** akkor kerül be az adatbányászási feladatok közé, amikor a hagyományos statisztikai idősor elemzési eszközök már nem alkalmazhatók a feladat bonyolultsága (túl sok változó) miatt.

A társításelemzés, a csoportosítás és az osztályozás esetén használható algoritmusokkal a 6.4, 6.5 és 6.6 fejezet foglalkozik.

Ezekkel a technikákkal elvben nagyon sok szabály, mintázat (minta) előállítható. Az előállítható minták jó része nem „érdekes” a felhasználók szempontjából. Azt szokás mondani, hogy egy minta érdekes, ha egyszerűen érthető, adott megbízhatósággal érvényes új, vagy kísérleti adatokon, hasznos és újszerű. A kapott eredmények érdekességi megítélésére **objektív** és **szubjektív érdekességi mértékek** adhatók meg. Pl. társítási szabályok megadása esetén objektív mérték lehet a kapott szabály megalapozottsága, gyakorisága, azaz az adatbázisban levő adatok (rekordok) százalékos aránya. Az objektív mértékeket mindig társítani célszerű a szubjektív mértékekkel, hiszen nem ugyanazok az eredmények lesznek érdekesek egy cég kereskedelmi igazgatója számára, mint annak az elemzőnek, aki az alkalmazottak teljesítményét vizsgálja. A szubjektív érdekességi mérték általában valamilyen felhasználói meggyőződést jelent. Ez alapján vannak a **nem az elvárásnak megfelelő** minták (felhasználói meggyőződésnek ellentmondó) és az **elvárásnak megfelelő** minták. Ez utóbbiak akkor érdekesek, ha megerősítik a felhasználó feltételezését.

Az érdekességi mértékeket kétféleképpen szokás figyelembe venni. Ha a mértékek figyelembevételére az adatbányászás után történik, akkor a feltárt mintákat érdekességük szerint rangsorolják, elhagyva az érdektelenekeket. Sokkal előnyösebb azonban, ha a mértékek magát az adatbányászási folyamatot irányítják, korlátozzák: ekkor eleve kizárják az érdekességi mértéknek eleget nem tevő minták keresését.

## 6.3. Adatbányászási módszertanok és lekérdező nyelvek

Az eddigiekből jól érzékelhető, hogy az adatbányászás egy összetett folyamat. Több lépésen át lehet eljutni az adatoktól a kinyert ismeretek hasznosításáig. Egyszerre igényel üzleti tudást, informatikai, matematikai és modellalkotási ismereteket.

Az adatbányászás alkalmazása a 90-es évektől vált intenzívebbé, és kezdődött el a speciális adatbányászási szoftverek kifejlesztése. Megszületett az első üzleti célú adatbányászási program a Clementine, amely az Integrál Solutions Ltd. terméke. Egyre nagyobb igény mutatkozott az adatbányászási feladatok végrehajtásához valamilyen módszertan kifejlesztésére. 1996-ban készült el az iparágaktól független adatbányászási módszertani szabvány a **CRISP\_DM (Cross Industry Standard Process for Data Mining)**, amelyet a Clementine gyártói és több iparág képviselői közösen hoztak létre. Ez a módszertan az adatbányászási folyamat mindegyik szakaszában ellenőrzési listákkal, útmutatásokkal, célok és feladatok meghatározásával segíti az elemzési feladat végrehajtását. Más módszertanok is ismertek, de a felmérések szerint a legelterjedtebb módszertan egyértelműen a CRISP\_DM (<http://www.crisp-dm.org/>).

A rugalmas és hatékony adatbányászási rendszerekkel szemben elvárás az ad hoc (eseti) és az interaktív adatbányászás támogatása. E tulajdonságok megvalósítására különböző adatbányászási lekérdező nyelveket fejlesztettek ki. Ilyen pl. a DMQL vagy a Microsoft OLE DB for Data Mining szabványosított adatbányászási nyelve.

A CRISP\_DM alapján egy adatbányászási folyamat életciklusa hat fázisból áll, amelyet a 6.3. ábra mutat. A nyilak mutatják, hogy az egyes fázisok között visszacsatolások lehetségesek, sőt bizonyos esetekben szükségesek is.

### 6.3.1. Az üzleti cél definiálása

Egy adatbányászási feladat sikeres megoldása szempontjából elengedhetetlen az üzleti cél pontos meghatározása, a korlátok és a lehetőségek felmérése.

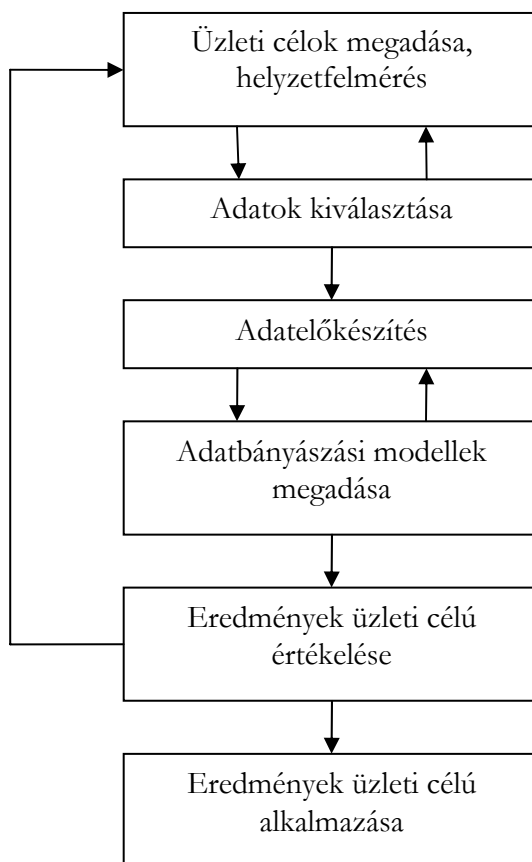
Ehhez egyrészt meg kell ismerni a hosszútávon elérendő cél(oka)t és elemzéssel meg kell adni azt is, hogy a cél(ok) milyen feltételek mellett érhető(k) el.

Másrészt el kell végezni a helyzetfelmérést: fel kell tární a rendelkezésre álló erőforrásokat (beleértve az idő- és anyagi erőforrást), megvizsgálni

a korlátokat, kockázatokat és lehetőleg meghatározni minden olyan egyéb tényezőt, amely befolyással lehet a feladat sikeres teljesítésére. Nem hagyható el a költségek és a várható haszon vizsgálata sem.

Ezek után következik a meghatározott üzleti cél „lefordítása” adatbányászási feladattá.

Ennek a fázisnak az utolsó lépéseként el kell készíteni az adatbányászási feladat részletes megvalósítási tervét, amely konkrétan tartalmazza az elemzési lépéseket az alkalmazandó technikákkal és eszközökkel együtt.



6.3. ábra: Az adatbányászás életciklusa a CRISP\_DM alapján

### 6.3.2. Az adatok kiválasztása

Ebben a fázisban kell megadni azt, hogy a megfogalmazott feladat megoldásához milyen adatokra van szükség. Itt általában egy iterációs szakasz

következik: összhangba kell hozni az elérendő célt és a rendelkezésre álló adatokat, ami az előző fázisba való többszöri visszalépéssel oldható meg.

A kiindulási adatok összegyűjtése saját adattárházból, információs rendszerekből, egyéni táblákból történhet, de szükség lehet külső, nem a vállalatnál keletkezett adatok beszerzésére is (pl. árfolyam adatok stb.).

Célszerű az összegyűjtött adatokat elemezni: megvizsgálni a tulajdonságaikat, alapstatisztikai számításokat végezni. Érdemes ellenőrizni az adatokat feltöltöttség, lefedettség és adathelyesség szempontjából is.

### 6.3.3. Adatelőkészítés

A rendelkezésre álló adatok előkészítése, előfeldolgozása azt a célt szolgálja, hogy javuljon az adatok minősége és ezzel az adatbányászás eredményessége. Ez a lépés igen fontos, mert megalapozott döntések csak megbízható adatok ismeretében hozhatók. Az adatokban fellépő anomáliák megkeresésére, kijavítására és szükség esetén az elemzendő adatmennyiség csökkentésére fordított idő bőven megtérülhet a döntéshozatal során.

Az adatelőkészítés műveleteit az alábbi, táblázatosan adott (6.4. ábra), folyóirat előfizetéseket tartalmazó rekordok esetén mutatjuk meg.

Ügyfélkód	Név	Cím	Előfiz. dátum	Folyóirat
1001	Kiss János	Petőfi u.5	06.01.15.	HVG
1001	Kiss János	Petőfi u.5	03.02.01.	Autós élet
1001	Kiss János		05.05.01.	Blikk
1023	Nagy József	Viola u.10.	01.01.01.	Blikk
1045	Nagy János	Rózsa u. 3.	02.03.01.	Sport újság
1056	Kiss János	Petőfi u.5	01.01.01.	Lakáskultúra

6.4. ábra: Kiindulási adatok

Az adatok előfeldolgozásánál leggyakrabban előforduló feladatok:

#### Adattisztítás

A rendelkezésre álló adatok gyakran hiányosak, zajosak, duplán előfordulhatnak, inkonzisztensek (belső ellentmondást tartalmaznak).

Hiányzó adatok esetén (lásd 6.4. ábra 3. sor) használható eljárás a sor törlése, a hiányzó érték manuális kitöltése, numerikus adatoknál az attribútum átlagértékkel történő pótlása (pl. ügyfél keresete adat esetén). Szokás az is, hogy a legvalószínűbb értéket használják a hiányzó érték pótlására,

aminek a meghatározása történhet regresszióval, vagy akár döntési fa létrehozásával is. Ez utóbbi módszer előnye az, hogy a legtöbb információt használja fel a hiányzó érték megbecslésére.

Az adatbázisban egy ügyfélnek több rekord is megfelelhet, de előfordulhat *duplázódás* tévedés miatt is. A példánkban a 6. rekord sejteti ezt az esetet. A duplázódások helyes megszüntetése a felhasználó döntése alapján, vagy mintaelemzési technikák segítségével történhet.

Az adatok esetében *zaj*nak az adatokra rakódott, véletlenszerű hibákat nevezzük. Ez a hiba igen gyakori, ha az adatok mérési eredményekből adódnak. A zaj miatt fellépő kiugró értékek általában távol vannak a helyes adatoktól. Az ilyen adatok javításának egyik módja az *adatsimítás*, amely a kiugró értéket a szomszédos adatok alapján módosítja. Használható módszerek továbbá a kosarazás és a kluszterezés is.

Az ellentmondásos adatok kiszűrésére már az adatok megadásánál lehetőség van. Az adatbáziskezelő rendszerek ugyanis lehetőséget nyújtanak arra, hogy a bennük tárolt információkra megszorításokat lehessen előírni.

### Adatok integrálása és transzformálása

Az **adatok integrálása** az a folyamat, amikor a kitűzött feladat megoldásához szükséges, különböző forrásból származó adatokat egyetlen adatárban egyesítik.

Az adatok integrálásakor megoldandó feladat az **egyedazonosítási probléma**. Eldöntendő kérdésként vetődhet fel például, hogy különböző helyről származó adatok esetében ugyanazt jelenti-e a *vevő\_kód* és a *vásárló\_ID*? Szükség lehet a **redundáns adatok** kiszűrésére, az **ellentmondó adatértékek felderítésére** és kijavítására.

Új adatok beszerzése és a felismert hibák javítása utáni állapotot a 6.5. ábra mutatja.

Az **adatok transzformálásával** a kiindulási adatokat az adatbányászás céljainak, az alkalmazni kívánt technikának megfelelő alakra hozzuk. Pl. a neuron hálózatok alkalmazása numerikus adatokat igényel, míg a döntési fáknál szimbolikus értékek is használhatók. A transzformációnál vigyázni kell arra, hogy ne történjen információvesztés.

Gyakran használt transzformáció a **normalizálás**, amikor az adatokat egy előre megadott tartományba vetítik le, pl. a [0,1] intervallumba. Összehasonlító módszereknél lényeges, hogy az egyes attribútumok azonos intervallumba essenek.

Ügyfélkód	Név	Szüil.dátum	Jövedelem	Gépkocsi	Cím	Előfiz. dátum	Folyóirat
1001	Kiss János	75-01-23	220 000 Ft	<i>van</i>	Petőfi u.5	06.01.15.	HVG
1001	Kiss János	75-01-23	220 000 Ft	<i>van</i>	Petőfi u.5	03.02.01.	Autós élet
1001	Kiss János	75-01-23	220 000 Ft	<i>van</i>	<i>Petőfi u.5</i>	05.05.01.	Blikk
1023	Nagy József	85-05-03	90 000 Ft	<i>van</i>	Viola u.10	01.01.01.	Blikk
1045	Nagy János	80-09-13	150 000 Ft	<i>nincs</i>	Rózsa u.3	02.03.01.	Sportújság
1001	Kiss János	75-01-23	220 000 Ft	<i>van</i>	Petőfi u.5	01.01.01.	Lakáskultúra

6.5. ábra: Bővített és javított adatok



## Az adatok redukálása

Az **adatredukció** célja, hogy az adathalmaz elemzése ne váljon lehetetlenné az elemzéshez szükséges idő miatt. Ilyenkor úgy kell az adathalmazt csökkenteni, hogy méretében jóval kisebb legyen, de a redukált adathalmazon végzett bányászat várhatóan az eredetivel azonos, vagy majdnem azonos eredményt adjon.

Lehetséges módszerek többek között a **dimenziócsökkentés**, az **adattömörítés** és a **számosságcsökkentés**. A dimenziócsökkentés úgy redukálja az adathalmaz méretét, hogy az adatbányászati feladat szempontjából érdektelen attribútumo(ka)t eltávolítja, vagy minden attribútumot figyelembe véve új, kevesebb számú attribútumot hoz létre. Az adatok tömörítése során az adatokat kódolással, vagy transzformációs műveletekkel hozzuk „tömörebb” alakra, míg a számosság csökkentésnél a tárolandó rekordok számát csökkentjük valamilyen elv alapján. Mindhárom esetben számos módszer létezik a redukció végrehajtására.

Példánknál a végleges adatokat a 6.6. ábra tartalmazza. Látható, hogy a címek helyett területi kódok szerepelnek, ezzel tömörítettük a túl részletesen megadott értékeket. Születési dátum helyett a kor szerepel, a jövedelmet is osztottuk 1000-el, így ez utóbbi két érték közel azonos nagyságrendű lett, így könnyebben értelmezhető és ábrázolható a jövedelmek és korosztályok közötti összefüggés. Ha nem annyira az idősorokra vagyunk kíváncsiak, hanem a különböző újságok olvasói közötti kapcsolatokra, akkor az előfizetési dátum nem fontos, ezért elhagyjuk.

Kód	Név	Kor	Jövedelem	Autó	Ter. kód	Folyóirat
1001	Kiss János	31	220	1	1	HVG
1001	Kiss János	31	220	1	1	Autós élet
1001	Kiss János	31	220	1	1	Blikk
1023	Nagy József	21	90	1	1	Blikk
1045	Nagy János	26	150	0	2	Sport újság
1001	Kiss János	31	220	1	1	Lakáskultúra

6.6. ábra: A végső adatok

### 6.3.4. Adatbányászási modellek megadása

Ebben a lépésben történik a megfelelően előkészített adatokból az új ismeret(ek) kinyerése, használva a rendelkezésre álló adatbányászási termékek algoritmusait.

A *megfelelő modellező technika kiválasztása* általában nem egyetlen algoritmus kiválasztását jelenti, mert az esetek többségében az algoritmusok alkalmazási területei átfedik egymást. Meghatározva az elemezni kívánt adatok típusát könnyebben tudunk dönteni, mert másféle algoritmus használható numerikus adatok esetén, vagy ha pl. jó adós, átlagos adós, rossz adós kategória változók szerint kell elemezni.

Meg kell határozni a *modell tesztelési módszert*: el kell dönteni, hogy a kapott eredmények helyességét, jóságát milyen módon mérjük. A rendelkezésre álló adatokból tanuló mintát és teszt mintát kell létrehozni. A tanuló minta segítségével létrehozható a modell, a teszt mintával pedig a modell pontosságát lehet mérni.

A *modellalkotás* a kiválasztott algoritmus alkalmazását jelenti. Az előző lépésben kiválasztott technikával elvégezzük a mintakeresést a rendelkezésre álló adatok között. E lépés eredménye nagyon erősen függ attól, hogy az elemző személy helyesen választotta-e ki az elemezendő adatokat és az alkalmazott technikát.

A *modell kiértékelése* a feltárt ismeretek értelmezését, megjelenítését jelenti. Ha az eredmények nem megfelelők, akkor az előző fázishoz való visszatéréssel kell folytatni a feladat megoldását.

### 6.3.5. A kapott eredmények üzleti célú értékelése

Ebben a fázisban olyan szempontból kell megvizsgálni a kapott eredményeket, hogy mennyire felelnek meg az előzetes üzleti elvárásoknak.

Célszerű a teljes elemzési folyamatot is áttekinteni, megvizsgálni, hogy az előzetesen megadott elemzési folyamat teljesen végrehajtásra került-e.

Dönteni kell a további teendőkről. Az adatbányászási feladat befejezése esetén az eredményeket ábrák, grafikonok, a modellezési technika hátterének ismerete nélkül is értelmezhető formában át kell adni a felhasználóknak. Ha az új ismeretek birtokában újabb kérdések vetődnek fel, vagy az eredmények a felhasználók számára nem jelentenek használható információkat, akkor a 6.3. ábra megfelelő iterációjával folytatódik a munka.

### 6.3.6. Az eredmények üzleti célú alkalmazása

A kapott eredmények alkalmazása azt jelenti, hogy az új információt be kell építeni valamilyen döntési folyamatba. Általában a döntések nem csak ezen információkra épülnek. Pl. egy reklám kampány esetén adatbányászási módszerekkel a potenciális ügyfeleket lehet kiválasztani, majd dönteni az alkalmazott reklám eszközökről. Ez a példa azt is érzékelteti, hogy az alkalmazott adatbányászás eredménye általában nem mérhető közvetlenül.

Ebben a fázisban kell megtervezni az eredmények felhasználását, beépítését az üzleti folyamatokba. Ha a vállalat mindennapi tevékenységét érintő alkalmazásról van szó, akkor folyamatosan biztosítani kell az alkalmazás „frissítését”, elkerülve a már nem aktuális adatokra épült modellek hosszú távú használatát. El kell készíteni a feladathoz tartozó összefoglaló jelentést, amely tartalmazza a végrehajtott lépéseket, a kapott eredményeket és a jövőre vonatkozó, más feladatoknál is használható tapasztalatokat. Célszerű ismertetni a pozitívumok mellett a negatívumokat is: hol van szükség felülvizsgálatra, módosításokra. Megismerve a felhasználók tapasztalatait, összehasonlítható a valóság és a modell. A különbségek feltárása új modellek megalkotását ösztönözheti, vagy az összehasonlítás akár olyan tény(eke)t is kimutathat, amely(ek) a továbbiakban üzleti célként is használható(k).

## 6.4. Társítási szabályok meghatározása

A társítási szabályok bányászatára tipikus példa a vásárlói kosár elemzése. Ilyenkor az a kérdés, hogy a vásárlók mely árukat vásárolják együtt, hiszen ez ötletet adhat pl. árucikkek boltokban egymás melletti elhelyezésére. Az ismertetésre kerülő alapfogalmak és algoritmusok bemutatására használt vásárlói kosarakat a 6.7. ábra tartalmazza.

TID transaction Identifier	Vásárolt termékek	Termékek kódja	Vásárolt termékek kódosan
1.	tej, kenyér, csoki, bor	tej → a	a,b,c,e
2.	tej, csoki, vaj	kenyér → b	a,c,d
3.	kenyér, bor	csoki → c	b,e
4.	kenyér, csoki, bor	vaj → d	b,c,e
5.	tej, bor	bor → e	a,e

6.7. ábra: Vásárlói kosarak, tranzakciók

### 6.4.1. Alapfogalmak

Jelölje  $T$  a tranzakciók (az adatbázis rekordok) halmazát, az  $E$  pedig az adatbázisban előforduló elemek halmazát. Példánkban  $T$  öt elemű, az  $E$  halmaz pedig  $E = \{a, b, c, d, e\}$ . Az  $E$  részhalmazait elemhalmazoknak, ha  $k$  elemet tartalmaz, akkor *k-elemhalmaznak* nevezzük.

Tekintsük az  $A = \{b, e\} = \{\text{kenyér}, \text{bor}\}$  2-elemhalmazt. Az elemhalmaz *előfordulási gyakorisága* azoknak a tranzakcióknak a száma, amelyek tartalmazzák az elemhalmazt. Példánkban  $A$  gyakorisága 3, mert az első, harmadik és negyedik tranzakció is tartalmazza a kenyér és bor elemeket. Gyakran százalékos formában használják a gyakoriságot, azaz az előfordulás és a tranzakciók számának hányadosaként. Az  $A$  elemhalmazt *gyakorinak* nevezzük, ha egy előre adott  $\sigma$  értékre fennáll, hogy

$$\text{gyakoriság}(A) \geq \sigma.$$

Az  $A$  és  $B$  halmazok közötti a társítási szabály egy

$$A \Rightarrow B$$

implikáció, ahol  $A \subseteq E$ ,  $B \subseteq E$  és  $A \cap B = \emptyset$ . Társítási szabályok keresésekor arra vagyunk kíváncsiak, hogy a  $B$  elemhalmazt a tranzakciók hány százaléka tartalmazza úgy, hogy az  $A$  elemhalmazt is tartalmazza. A keresett érték a  $P(B|A)$  feltételes valószínűség, amelyet a társítási szabály *bizonyosságának* neveznek és a következőképpen számítható:

$$\text{bizonyosság}(A \Rightarrow B) = \frac{\text{gyakoriság}(A \cup B)}{\text{gyakoriság}(A)}.$$

A társítási szabályokra előírt  $\gamma$  küszöbértéket minimális bizonyosságnak nevezik.

Egy szabály akkor *érvényes társítási szabály*, ha egyszerre teljesíti a minimális gyakoriság és a minimális bizonyosság követelményét:

$$\text{gyakoriság}(A) \geq \sigma \text{ és } \text{bizonyosság}(A \Rightarrow B) \geq \gamma.$$

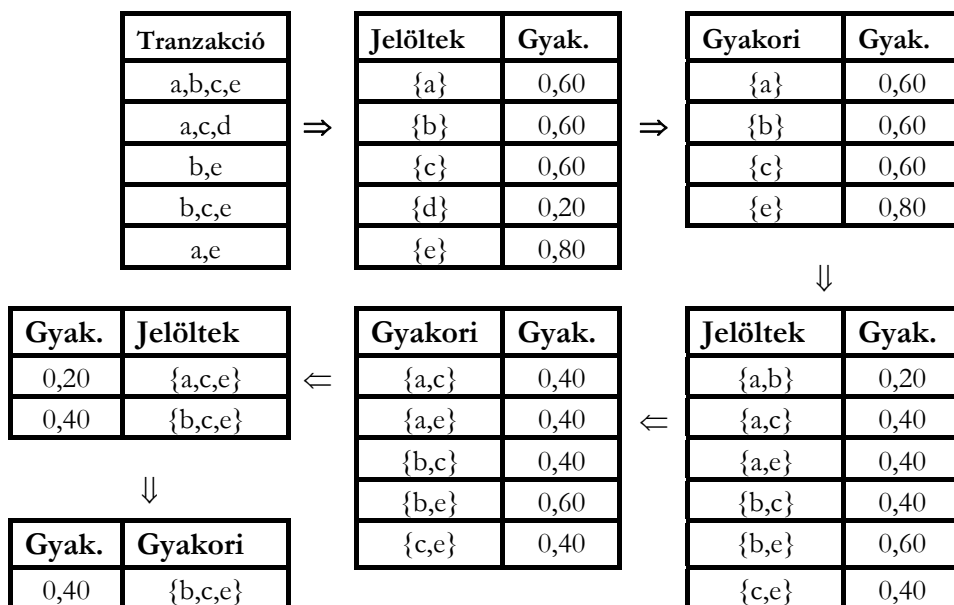
A társítási eljárások meghatározása két lépésben történik: meg kell keresni a gyakori elemhalmazokat, majd ezekből kell előállítani az érvényes társítási szabályokat.

### 6.4.2. Gyakori elemhalmazok keresése Apriori algoritmussal

Az algoritmus a nevét onnan kapta, hogy a gyakori elemhalmazokra vonatkozó előzetes (*a priori*) ismereteket használja fel. Nevezetesen azt, hogy ha egy elemhalmaz nem gyakori, akkor egy elemmel bővítve ismét nem gyakori elemhalmaz adódik. Ez fordítva is igaz: egy gyakori elemhalmaz minden rész elemhalmaza is gyakori.

A szintenként haladó (a kisebb méretű elemhalmaztól a legnagyobb méretűig haladó) módszer röviden: először a gyakori 1-elemhalmazokat kell megkeresni ( $L_1$  halmaz),  $L_1$ -t használva következik  $L_2$  (gyakori 2-elemhalmazok összessége) meghatározása, majd ennek a segítségével  $L_3$  meghatározása, és így tovább, amíg már nem található újabb gyakori  $k$ -elemhalmaz.

A 6.8. ábra egy szintenként haladó Apriori algoritmus illusztrálása látható. Minden lépésben két feladat van: a jelölt generálás és a gyakoriság ellenőrzése. A gyakoriságot nem teljesítő elemhalmazokat nem vizsgáljuk tovább.



6.8. ábra: Az Apriori algoritmus 30 %-os gyakorisági küszöbvel

A gyakori  $k$ -elemhalmazokból a jelölt  $k+1$ -elemhalmazok illesztéssel határozhatók meg. Két gyakori  $k$ -elemhalmaz összeilleszthető, ha lexikografi-

kusan rendezett első  $k-1$  elemük megegyezik és az utolsó különbözik. Pl. ha  $\{b,c,e\}$  és  $\{b,c,d\}$  gyakori elemhalmazok lennének, mivel összeilleszthetők, az összeillesztés után az új jelölt a  $\{b,c,d,e\}$  4-elemhalmaz lenne. Az algoritmus véget ér, ha nem lehet új jelöltet állítani, vagy a vizsgált jelöltek közül egyik sem éri el az adott gyakorisági küszöböt.

### 6.4.3. Társítási szabályok generálása gyakori elemhalmazokból

Az érvényes társítási szabályok előállítására két lépésben történik. Először minden gyakori  $Z$  elemhalmazt fel kell bontani az összes lehetséges  $X, Y$ , két diszjunkt és nem üres halmaz párára. A kapott halmaz párok közül azok lesznek érvényesek, amelyek teljesítik a minimális bizonyossági követelményt. (Már csak ezt kell vizsgálni, mert a gyakori elemhalmazokra fenn áll a gyakorisági követelmény.)

Példánknál a lehetséges társítási szabályok:

$\{b,c\} \Rightarrow \{e\}$	bizonyosság: $2/2=100\%$
$\{b,e\} \Rightarrow \{c\}$	bizonyosság: $2/3=66\%$
$\{c,e\} \Rightarrow \{b\}$	bizonyosság: $2/2=100\%$
$\{b\} \Rightarrow \{c,e\}$	bizonyosság: $2/3=66\%$
$\{c\} \Rightarrow \{b,e\}$	bizonyosság: $2/3=66\%$
$\{e\} \Rightarrow \{b,c\}$	bizonyosság: $2/4=50\%$

Azt mondhatjuk, hogy ha 70% a bizonyossági küszöb, akkor csak az első és harmadik szabály érvényes, azaz  $\{\text{kenyér, csoki}\} \Rightarrow \{\text{bor}\}$  és  $\{\text{csoki, bor}\} \Rightarrow \{\text{kenyér}\}$ .

### 6.4.4. További lehetőségek

Az ismertetettől kívül számos módszer létezik társítási szabályok feltárására. Más módszer használatos, ha mennyiségi társítási szabályt keresünk (pl. életkor és jövedelem között), vagy ha többdimenziós szabályt akarunk feltárni. (Az előzőleg tárgyalt példa egydimenziós volt.)

A szabálykereső eljárások más és más módon kereshetik a jelölteket: szintenként haladva (pl. ismertetett eljárás), vagy mélységben haladva (ECLAT és FP-growth algoritmus). Eltérő lehet a jelöltek gyakoriságának számlálási technikája is.

Fontos kérdés a társítási szabályok esetében az ún. érdekességi mutatók használata, amelyek segítségével a szabályok rangsorba állíthatók különlegességük, használhatóságuk alapján, vagy szűrhetők, a számuk csök-

kenthető, ha túl sok adódik. A használatos objektív mutatók közé tartozik pl. a Lift mutató, vagy a kovariancia analízisnél használatos RI (Rule Interest) mutató.

## 6.5. Csoportosítási (kluszterelemzési) algoritmusok

A csoport, vagy kluszter olyan adatelemek halmaza, amelyekre fennáll, hogy az azonos csoporton belüli adatelemek hasonlóak egymáshoz, míg a különböző csoportokhoz tartozók különböznek egymástól az adatbázis attribútumaival kifejezve. A csoportosítás a mindennapi életben igen gyakran előforduló feladat. Legismertebb alkalmazása piackutatási célból a vásárlói csoportok meghatározása, de csoportosítás segíthet a gépjármű biztosítással rendelkezők közül a nagy kárigényű ügyfelek csoportjainak azonosításában is. Fontos alkalmazás az is, amikor a kluszterelemzést előkészítő lépésként használják egyes algoritmusoknál, lásd. 6.3.3 fejezet.

### 6.5.1. Csoportosítási algoritmusokkal szembeni követelmények

Az adatbányászási környezetben alkalmazott algoritmusokkal szembeni elvárások közül az egyik legfontosabb a *skálázhatóság*, a nagy adathalmazon való hatékony csoportképzés. A nagy adatbázisokban ugyanis nem ritka a több millió adatelem előfordulása sem.

Fontos, hogy az algoritmusok *különböző adattípusok* esetén is alkalmazhatók legyenek. Az algoritmusok egy része numerikus adattípusok kluszterezésére alkalmas. A megoldandó feladatok azonban gyakran megkívánják egyéb adattípus (bináris, felsorolás típusú, ezek valamilyen keveréke) feldolgozását is. Emellett az is elvárás, hogy az algoritmus ugyanazokat a csoportokat hozza létre, ha más sorrendben kapja az adatrekordokat, és a zajos, hibás adatokat is kezelni tudja.

Lényeges az is, hogy az algoritmus *minimális felhasználói közreműködést* igényeljen: ne kelljen olyan paramétereket megadni, amelyeket csak becsülni lehet. Ez probléma a felhasználónak is, de nehézkessé teszi a kapott csoportosítás jóságának ellenőrzését.

A valós élet dolgaihoz való alkalmazkodás teszi szükségessé, hogy lehetőség legyen az előzetes ismeretek alapján adódott *korlátozó feltételek érvényesíthetőségére*.

A kluszterezés végeredményét olyan formában kell előállítani, hogy értelmezhető, áttekinthető és felhasználható legyen.

### 6.5.2. Adatok hasonlósága, különbözősége

A csoportokat, klusztereket úgy kell létrehozni, hogy a hasonló adatelemek, objektumok kerüljenek egy csoportba. Ezért szükség van az adatok hasonlóságát, vagy különbözőségét valamilyen mértékkel meghatározni.

Ha a vizsgált objektumok  $n$  tulajdonsággal (attribútummal) rendelkeznek, akkor  $N$  objektum esetén a teljes objektum halmaz megadható az  $N \times n$  méretű *adatmátrix* segítségével:

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nn} \end{bmatrix},$$

ahol soronként az egyes objektumok tulajdonságai találhatók. A *különbözőségi mátrix* az elemek közötti különbség (távolság) ábrázolására szolgál, amely egy  $N \times N$  méretű mátrix, amelyben  $d(i,j)$  adja meg az  $i$  és  $j$  elemek közötti különbözőséget. A  $d(i,j) \geq 0$ , és ha  $i$  és  $j$  nagyon hasonló, akkor közel van 0-hoz, és annál nagyobb lesz  $d(i,j)$  értéke minél jobban különböznek egymástól  $i$  és  $j$ . Mivel  $d(i,j) = d(j,i)$  és  $d(i,i) = 0$ , ezért a különbözőségi mátrix a következő felépítésű lesz:

$$\begin{bmatrix} 0 & & & & & \\ d(2,1) & 0 & & & & \\ d(3,1) & d(3,2) & 0 & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ d(N,1) & d(N,2) & \cdots & d(N,N-1) & 0 & \end{bmatrix}.$$

Ha a hasonlósági értékekre van szükség, azt is egy hasonló felépítésű mátrixban, a *hasonlósági mátrixban* lehet megadni, de ott 1 áll a főátlóban, és minél kisebb  $d(i,j)$  értéke, annál jobban különböznek egymástól  $i$  és  $j$  objektumok.

Az egyes objektumok közötti hasonlóság(közelség)/különbözőség kiszámítása különböző típusú attribútum értékek esetén más és más elvek alapján történik.

**Folytonos értékek** esetén a közelséget leggyakrabban valamilyen távolsággal szokás megadni. (A távolság egy különbözőségi mérték.) Ilyenkor az objektumokat  $n$ -dimenziós adatvektornak tekintve, az  $n$ -dimenziós



térben legelterjedtebb távolság mértéket, az Euklideszi távolságot alkalmazzák. De használható a Manhattan-távolság, vagy a Minkovszki-távolság is. Ha az egyes tulajdonságok értéke nem azonos skálán mozog a többivel, akkor ezek domináns szerepet játszhatnak a távolság számításában. Ilyenkor standardizálással célszerű azonos intervallumskálára leképezni őket még a távolság számítása előtt.

**Bináris változóknak** csak két értéke lehet: 0 vagy 1. Ebben az esetben az objektumok hasonlósági mértékét határozzák meg, amely általában  $[0,1]$  intervallumba eső érték. Bináris változók esetén előállítható a *kontingencia* táblázat, amelyből számolható az *egyezőségi együttható*, az objektumok hasonlóságát kifejező érték. Ez az érték más összefüggéssel határozható meg szimmetrikus (egyik érték, sem a 0, sem az 1 nem bír kitüntetett szereppel) és aszimmetrikus (az egyik érték kiemelt fontosságú) bináris változóknál.

A **felsorolás típusú** változóknál a különbözőséget szokás számolni az ún. *egyszerű egyezőség* elve alapján.

**Vegyes típusú változók** esetében, amikor az objektumokat jellemző tulajdonságok nem azonos típusúak, a fent említett hasonlósági/különbözőségi mértékek kombinációját szokás használni.

### 6.5.3. A k-átlag módszer

A k-átlag módszer a klusztereket partícionálással meghatározó algoritmusok közé tartozik. A partícionáló módszerek célja  $N$  objektum  $k$  számú csoportba sorolása ( $k$  értéke előre adott), jellemzője pedig, hogy a vizsgált objektumoknak egy lehetséges csoportosítását adják eredményül, amelyet egy iterációs folyamatban határoznak meg.

A partícionáló módszerek alkalmazásánál az egyik fontos kérdés a  $k$  csoport szám meghatározása, ami általában a felhasználó feladata. Ha nem áll rendelkezésre megbízható érték, akkor szokás  $k$  különböző értékeire elvégezni a csoportosítást és a leginkább hasznosíthatónak látszó eredményt fogadják el végeredményként. Másik fontos kérdés, hogy az algoritmusok mi alapján döntsék el az egyes objektumokról, hogy melyik csoportba tartozzanak. Az objektumok csoportba sorolására leggyakrabban az ún. négyzetes hiba (egy objektumnak a csoport középpontjától való távolságának a négyzete) vizsgálatát használják.

A k-átlag algoritmus lépései adott  $N$  objektum és  $k$  csoport szám esetén következők:

1. A  $k$  darab elem véletlenszerű kiválasztása, ezek lesznek kezdetben a  $k$  darab csoport középpontjai (átlagai).
2. A fennmaradó elemek mindegyikének hozzárendelése ahhoz a csoporthoz, amelyiknek a középpontjához a legközelebb van. (A négyzetes hiba a legkisebb.)
3. Meg kell határozni a kialakult kluszterek új középpontját (átlagát).
4. A 2. és 3. lépés ismétlése mindaddig, amíg a 2. lépésben már egyetlen elem sem kerül új csoportba.

Más megállási kritérium is létezik. Gyakran akkor fejeződik be az iteráció, ha a csoportosítás teljes négyzetes hibája (az egyes elemek négyzetes hibáinak összege) egy előírt értéknél kisebb változást mutat két egymás utáni lépésben.

Az algoritmus hátránya, hogy érzékeny a kiindulási adatokra, akkor ad jó eredményt, ha tömör, jól elkülönülő csoportokat kell meghatározni. További hátrány még, hogy csak vektortérben jellemezhető adatok esetén használható. Az algoritmus továbbfejlesztett változata, a  $k$ -medoid algoritmus már képes tetszőleges tulajdonságokkal rendelkező adatokat használni.

#### 6.5.4. További lehetőségek

A csoportok meghatározására további módszerek állnak rendelkezésre. A **hierarchikus módszerek** a vizsgált objektumok nem egy, hanem több lehetséges csoportosítását adják eredményül. A hierarchikus módszerek két módon hozzák létre a csoportokat: vannak az *egyesítő* és a *felosztó* módszerek. Az egyesítő módszerek általános elve az, hogy kezdetben minden objektumot külön csoportban helyeznek el, majd lépésenként egyesítik az egymáshoz legközelebbi csoportokat (minden lépésben csak egy egyesítés történik). Addig folytatódik az egyesítés, amíg minden objektum egy csoportba kerül. A felosztó módszer ennek a fordítottja, ott kezdetben minden objektum egy csoportban van, amelyet lépésenként újabb csoportokra bontunk mindaddig, amíg minden elem külön csoportba nem kerül. Ezek az algoritmusok általában hasonlósági mátrixok alapján dolgoznak. Előnyük, hogy tetszőleges attribútumok esetén használhatók.

A **fuzzy módszerek** a kialakult halmazokat fuzzy halmazként kezelik. A fuzzy halmazok esetében egy elem több halmaznak is része lehet, mégpedig a különböző halmazoknak *különböző mértékben* lehet a tagja, ezért a fuzzy csoportosító módszerek alkalmazása nem elkülönült halmazokat eredményez. E tulajdonsága miatt nevezik ezeket a módszereket lágy cso-

portosító módszereknek, ellentétben a kemény csoportosító módszerekkel, ahol a kialakult csoportok mindig egyértelműen elkülönülnek egymástól.

Csoportok képzésére alkalmasak még a **neuronháló módszerek**, vagy a különböző **sűrűség alapú módszerek** is.

## 6.6. Osztályozási algoritmusok

Az osztályozás az adatbányászás egyik leggyakrabban használt formája. Az osztályozás az osztályozási szabály „megtanulását jelenti”, amely szabály alkalmas arra, hogy ismert változók (osztálycímke) alapján az adatbázis elemeit, a mintákat osztályokba sorolja. (Osztályozásnál szokás az adatrekordokat mintának nevezni.) A szabályok alkalmazása lehetővé teszi új minták osztályba sorolását is.

### 6.6.1. Az osztályozás lépései

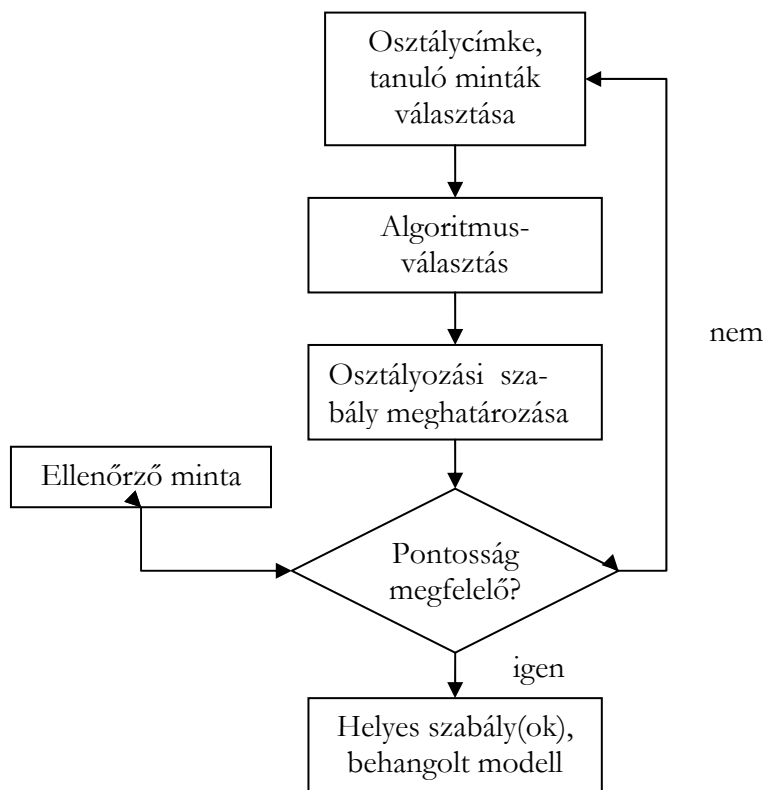
Az osztályozási módszerek alapvetően a következő két lépésből állnak: az első lépés az osztályozási szabály megtanítása (modellkészítés) és a modell pontosságának ellenőrzése, a második lépés pedig a modell felhasználása új adatminta ismeretlen osztálycímkéjének meghatározására.

Az első lépés folyamatát a 6.9. ábra mutatja. Először meg kell határozni az *osztálycímké*t, azaz ki kell kiválasztani egy attribútumot, amely alapján az osztályozás történik. Ennek olyannak kell lennie, amelytől a többi ismérv függ, mert különben nem lenne lehetséges az új minták ismeretlen osztálycímkéjének előrejelzése. Ezután a *tanuló minták* meghatározása következik: a rendelkezésre álló ismert osztálycímkéjű minták közül választunk. Ezekkel a mintákkal történik a választott algoritmus szerint az osztályozási modell előállítás. A modellek általában *ha-akkor* típusú osztályozási szabályként, döntési fák, vagy matematikai formulák alakjában adhatók meg.

A modell pontosságának mérése igen fontos lépés, amely a teszt minták segítségével történik. Ismert osztályozási címkejű minták esetén azt kell megvizsgálni, hogy az egyes mintákat az osztályozási szabály helyes osztályokba sorolja-e, vagyis hogy a minta valódi osztálycímke értéke megegyezik-e a számított osztálycímke értékkel. A modell annál pontosabb, minél nagyobb a helyesen osztályozott teszt minták aránya.

A teszt minták kiválasztása többféleképpen történhet. Egyik mód, amikor a rendelkezésre álló ismert osztálycímkejű mintákat véletlenszerűen két diszjunkt halmazra osztják, és az egyik lesz a tanuló minták halmaza, a másik pedig a teszt mintáké. Ilyenkor általában a tanuló minták szá-

ma nagyobb, szokásos arány a 2:1. További teszt minta megadási technika a *kereszt-validálás*, a *leave-one-out*, vagy a *bootstrap*.



**6.9. ábra:** Az osztályozás első lépése

A tanulás és a tesztelés iteratív folyamat, amelynek a végén az alkalmazás szempontjából megfelelő pontosságú modellt megkapjuk. Ezt a modellt *behangolt modellnek* szokták nevezni.

A második lépés a kapott modell, osztályozási szabály előrejelzésre használata. Ez új adatminták esetén a minta ismeretlen osztálycímkéjének meghatározását jelenti a többi, ismert attribútum értékének függvényében.

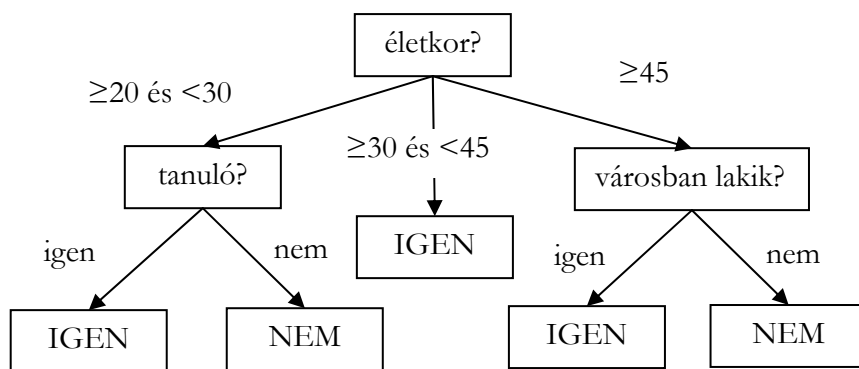
Osztályozás segítségével csak diszkrét érték előrejelzése lehetséges (az osztálycímke diszkrét értéke miatt), ezért folytonos érték előrebecslésére más módszert, valamilyen regressziós módszert kell használni.

### 6.6.2. Osztályozás döntési fa segítségével

Az osztályozási feladatok esetében a leggyakrabban alkalmazott eszköz a döntési fa, amely egy fa formájú folyamatábra. A fa gyökeréből indulva köztes csomópontokon keresztül lehet eljutni a fa leveleiig. Minden csomóponthoz egy attribútumra vonatkozó kérdés tartozik. A kérdésekre adott válaszok segítségével a fa valamelyik leveléhez jutunk, amelyek az egyes osztályok címkeit tartalmazzák.

A döntési fa egy szabálybázist határoz meg (a döntési fából kinyerhető szabályok összességét), amelynek mindegyik szabályát a fa gyökeréből indulva, egy levél felé haladó útvonal alapján lehet előállítani. A „ha-akkor” típusú szabály feltételi részét az útvonalba eső csomópontokhoz tartozó feltételek ÉS művelettel való összekapcsolásával kapjuk, míg a szabály kimenetelét az útvonal végén levő levél adja meg.

A 6.10. ábra egy döntési fa látható, ami egy kereskedelmi cég eladásai alapján azt tudja megbecsülni, hogy egy vásárló vesz, vagy nem vesz számítógépet. Erről a döntési fáról öt szabály olvasható le (öt levél található rajta).



6.10. ábra: Döntési fa számítógép vásárlás vizsgálatára

Egy szabály pl. az, hogy aki 30 év alatti és még tanul, az vesz számítógépet (szüksége van rá). De egy olyan szabály is megadható, hogy aki idősebb 45 évesnél és nem városban lakik, nem vesz számítógépet (nem fogékony az új dolgokra). Jól látszik a döntési fának az az előnye, hogy könnyű az osztályozási szabályok kiolvasása és alkalmazása.

A döntési fák elkészítése felülről lefelé történik. Kezdetben minden minta a fa gyökeréhez tartozik. A továbbiakban minden lépésben egy kiválasztott attribútum alapján elágazásokat (csomópontokat) készítünk, annyit, ahány diszkrét értéke van az attribútumnak. A mintákat ezek szerint osztjuk szét diszjunkt részhalmazokra. A rekurzív felosztást (az előző felosztás eredménye lesz a kiindulás a következő lépéshez) addig végezzük, amíg az egy csomópontoz tartozó összes minta azonos osztályból való, vagy már nincs olyan attribútum ami alapján további bontás készíthető. Az is az algoritmus végét jelentheti, ha már az összes mintát figyelembe vettük. Ezekben az esetekben a csomópontból levélcúcs lesz.

A döntési fa elkészítésénél többféle módon lehetséges annak a változónak, attribútumnak a kiválasztása, amely alapján a minták felosztása történik. Egyik lehetőség az *információnyereség* vizsgálata. Az ezt használó algoritmusok a döntési fa egyes csomópontjaiban minden lehetséges attribútumra meghatározzák az információnyereség értékét, és a maximális információnyereséggel rendelkezőt választják az adott csomópontoz tartozónak.

A döntési fák elkészítése után szokásos lépés a *döntési fák tisztítása*. Erre azért van szükség, mert a döntési fák túlzottan illeszkedhetnek az adatokhoz, nagymértékben tükrözhetik a tanuló minták hibáit, „zajait”. Ilyenkor a legkevésbé megbízható ágak eltávolítása történik meg. *Előmetszés* esetén hamarabb leállítják a fa felépítését, *utómetszésnél* pedig a kész döntési fáról távolítják el a nem kívánatos részeket. Mindkét esetben megfelelő mértékek állnak rendelkezésre a tisztításhoz.

### 6.6.3. További lehetőségek

A gyakorlatban jól használhatók a különféle *Bayes-osztályozók*, amelyek a valószínűségszámításból ismert Bayes-tételen alapulnak. A naiv Bayes-osztályozó segítségével megbecsülhető, hogy egy adott minta milyen valószínűséggel tartozik valamelyik osztályba. (A naiv jelző arra a naiv feltételezésre utal, hogy az attribútumok függetlenek egymástól.) Jól használható a minták és az attribútumok nagy számú eseténél, de még hiányos adatoknál is.

A *k-legközelebbi szomszéd* technika feltételezi, hogy az adatok numerikusak, így minden tanuló mintát egy  $n$  dimenziós tér egy pontjának lehet tekinteni. Az alapötlet az, hogy meg kell keresni a  $k$  darab legközelebbi szomszédot, és az ismeretlen minta abba az osztályba fog tartozni, amely a  $k$  szomszéd között a leggyakoribb.

Osztályozni lehet még *neuronhálók* segítségével, vagy genetikus algoritmusokkal is.

## 6.7. Adatbányászási szoftverek

Adatbányászási feladatokra alkalmas szoftverek forgalmazásával Magyarországon sok cég foglalkozik. A széles választékot mutatja az alábbi táblázat is. (Az adatok a [www.adatbanyasz.hu](http://www.adatbanyasz.hu) weblapról származnak.)

Termék	Magyarországi forgalmazó	URL
Clementine	SPSS Hungary	<a href="http://www.spss.hu">www.spss.hu</a>
Enterprise Miner	SAS Institute	<a href="http://www.sas.com/hungary">www.sas.com/hungary</a>
Intelligent Miner	IBM Magyarország	<a href="http://www.ibm.hu">www.ibm.hu</a>
Darwin	Oracle	<a href="http://www.oracle.hu">www.oracle.hu</a>
Datascopé	Cygron	<a href="http://www.cygron.hu">www.cygron.hu</a>
MineSet	Silicon Computers Kft.	<a href="http://www.silicon.hu">www.silicon.hu</a>
Scenario	Axis	<a href="http://www.axis.hu">www.axis.hu</a>

6.11. ábra: Magyarországon forgalmazott adatbányászási termékek

A fenti termékekre általában jellemző, hogy képesek az adatbányászási módszertan minden lépésének a támogatására. Ha a termék nem tartalmazza az adatelőkészítés funkciót, akkor elterjedt megoldás, hogy az adatbáziskezelőhöz integrálják az adatbányászási eszközöket. Így működik például az SAS Enterprise Miner szoftvere, vagy az MS SQL Server.

A legrégebben és legelterjedtebben használt adatbányászási szoftver, a Clementine az összes nagy adatbázissal összekapcsolható. Az adatelőkészítés és az elemzés is hatékony vizuális kezelői felületen történik. Rugalmasságát mutatja, hogy egyszerűen illeszthetők hozzá saját fejlesztésű algoritmusok is.

Az adatbányászás megismerését, tanulmányozását nagyban elősegítheti a mindenki számára ingyenesen hozzáférhető WEKA (Waikato Environment for Knowledge Analysis) programcsomag, amelyet az Új-Zélandon működő Waikato Egyetemen fejlesztettek ki. A moduláris felépítésű program Linux vagy Windows környezetben is használható. Részletes információ a <http://www.cs.waikato.ac.nz/m1/weka> honlapon található.

## 7. A Comshare rendszer

### 7.1. A termék első generációjának kialakítása

A Comshare Corporation vezető szoftverfejlesztő és -forgalmazó világ cég. Központja az Egyesült Államokban van. A Comshare cég fejlesztései kizárólag a vezetői döntéstámogatásra, a vezetői információs rendszerekre irányulnak. Mintegy 35-éves tapasztalat áll e tevékenységük mögött. Világpiaci részesedésük a VIR-kategóriában a legelső.

Az 1980-as években, ill. az 1990-es évek elején a vezető szoftvercégek a termékeiket elsősorban az IBM nagygépi változataira, ill. a DEC-nek a VAX gépeire szánták. Ugyanezen az úton indult el a Comshare is. A későbbiekben a VIR-termékeket fokozatosan kiterjesztették a személyi számítógépes piacra is. Ma már döntően ezen az utóbbi platformon működnek a rendszereik. Itt fontos megjegyezni, hogy a PC-vonal kiépülése lehetővé tette, hogy a VIR-ek kisebb vállalatoknál (kisebb teljesítmény, lényegesen alacsonyabb ár) is teret nyerjenek, ezen túl pedig alsóbb vezetői szintekhez is eljussanak.

A Comshare két termékkel kezdte el a piaci szereplését. Ezek: a *Comshare Commander* és a *Comshare Quickstart*. Az előbbi a teljes kiépítésű VIR-t képviseli, míg az utóbbi a viszonylag gyors, kísérleti jellegű elindulást segíti elő a menedzsment számára.

A *Quickstart* lehetővé teszi, hogy a cég egy prototípus VIR-t fejlesszen ki, viszonylag alacsony költséggel. Ebben fel tudja használni a PC-ken keletkező adatokat is, a saját lokális hálózatának bevonásával. A *Quickstart* igen jó arra, hogy egy kezdeti VIR-t hozzanak létre vele, azt elemezzék, és a tapasztalatok alapján már a *Commander* felhasználásával építsék ki a végleges megoldást. A másik irányú felhasználás egy-egy kisebb igényű, kisebb teljesítményű döntéstámogató eszköz kialakítása, ami lehet önálló, egy adott részterület kiszolgálására, de szatellitként is csatlakozhat a vállalat elsődleges, fő VIR-jéhez.

A következő alpontban részletesebben ismertetjük a *Commander* működési jellemzőit. Mint látni fogjuk, az egyes funkciók jól tükrözik azokat a lehetőségeket és használati elveket, amelyek kifejezetten a VIR-ekben nyilvánulnak meg.



## 7.2. A Commander funkciói

A Commander főmenüje közvetlen intuitív elérést biztosít a társaság stratégiai információihoz. Használatához nem kell sem külön tanfolyam, sem dokumentáció. A főmenüből elérhető funkciók, ill. rendszerkomponensek a következők:

### 7.2.1. Beszámoló jelentés kezelése

Az írásos beszámoló jelentések tömör és informatív megjelenítését a *Briefing Book* nevű komponens teszi lehetővé. A jelentések hierarchikus rendben érhetőek el. Tartalmaznak szöveget, ábrákat, táblázatokat, valamint diagramokat. Ha a felhasználó rámutat egy kiemelt szövegrészre vagy egy diagramra, azonnal megjelenik a képernyőn az ezekhez tartozó háttér információ. A háttér információ lehet pontosan olyan, mint egy írásbeli jelentés, de kiegészíthető színes diagramokkal, melyekben a jelentés időtartamánál hosszabbra nyúló trendek jeleníthetők meg grafikusán.

A Commander Briefing Book jobb tájékoztatást nyújt, mint egy írásos beszámoló jelentés, a következők miatt:

- Minden vezető testreszabott jelentésekkel és diagramokkal dolgozhat, csökkentheti a képernyőn megjelenő, de nem fontos információ mennyiségét.
- Színekkel meg lehet jelölni és kiemelni azokat az eltéréseket, amelyeknek a nagysága elér valamely előre megadott értéket.
- Az adatok bármikor karbantarthatók. Az információ azonnal továbbítható.
- Váratlan vagy sürgős ügyek intézésekor egy elektronikus irat kikeresése gyorsabb, egyszerűbb és kényelmesebb. Számítógépes adatelérés esetében nem kell becsült adatokkal dolgozni, mert a pontos adatok késedelem nélkül rendelkezésre állnak. Ez gyorsítja a döntési folyamatot, javítja a döntések minőségét, és csökkenti a döntések kockázatát.

A Briefing Book kényelmesen használható PC-ken, hálózati kapcsolatban. Használata egy VIR folyamat első fázisa is lehet. A legtöbb vezető ugyanis nem elégszik meg egyszerű táblázatos jelentésekkel, az információkat szöveggörnyezetben kívánja látni. Ha a vezető például a havi bérfizetés adatait egy hosszabb folyamat részeként vizsgálja, megtudja, hogy a kifizetett bértömeg időben növekszik-e vagy csökken. Hasznos információkat lehet nyerni a tervezett és tényleges költségek, továbbá az előrejelzések összeha-

sonlításával nyerni. Összevethetők bizonyos vállalati adatok a versenytársak nyilvánosságra kerülő hasonló adataival.

A tervadatok és a tényadatok közötti eltérés a színjelölésekkel könnyen észrevehető. Egy olyan jelentés, amiben azok az adatok vannak kiemelve, melyekben a terv- és ténytérvek eltérése nagyobb egy megadott értéknél, gyorsabban áttekinthető, mint egy hagyományos jelentés. Ez a számítógépes feldolgozás egyik előnye. Az alkalmazó minden egyes jelentés esetében egyedileg megadhatja az általa elfogadhatónak tartott eltéréstartományt. Ha egy jelentés megjelenik a képernyőn, azok az adatok, melyek nem esnek a megengedett eltéréstartományba, színesen kiemelve láthatók. A kedvezőtlen irányú eltérések vörös, míg a kedvezőek zöld színűek.

Ezekben a jelentésekben a működési adatokat a szokásos nyereség és veszteség forma mellett egyedi módon is meg lehet jeleníteni, például üzletrészek, gyártósorok, kereskedelmi csatornák, körzetek, felhasználói csoportok stb. szerinti bontásban. Ilyen adatok írásos jelentésekben nem szoktak szerepelni, mert a jelentést használhatatlanul bőbeszédűvé tennék, még ha tömörített formában lennének is. A nem szokványos tartalmú táblázatok világíthatják meg legjobban azokat a problémákat, melyek egy vezetőtől beavatkozást igényelnek. Ugyanakkor számítógép esetén tetszés szerint összeállított táblázatok lehet egymás után megjeleníteni.

Ha egy vezető számítógépes eltérés-kijelzést használ, hamarabb felfedezheti a problematikus helyeket, s ezzel nő a sikeres beavatkozás esélye. Mivel minden felhasználó maga állítja be az értéktartományokat, ezért nagyobb bizalommal használ kisebb információ-mennyiséget is. Mindezt a saját feladat-megoldási módszeréhez is hozzá tudja igazítani.

További módszer lehet jelentéseknek szöveggel való kiegészítésére előre megadott vizsgálat utak használata. Menüszerűen kiválaszthatóvá tehető a Briefing Book jelentéseinek önálló tételei. Ha egy magasabb szintű jelentés valamelyik adatára rámutatva alacsonyabb szinten található részletek, vagy más szempontok szerinti magyarázatok hívhatók elő, a felhasználó azt érezheti, hogy az információhalmazban úgy mozog, mint egy fényképező a természetben a zoom optikájával.

Egy vörös színű eltérés esetében, rámutatva a vörös számra, megjeleníthető egy másik jelentés, amiben az eltérés okára vonatkozó részletek olvashatók. Ez a zoomolási folyamat az egyik jelentésből vagy táblázatból kiindulva egy másikban folytatható. A vezető így jobban megismeri a részleteket, ill. a problémát más módon is tanulmányozhatja. Például a költségelőirányzat-költségfelhasználásra vonatkozó eltérési jelentésből át lehet

lépni egy trendeket felmutató jelentésbe, ahol havi, negyedéves vagy éves bontásban szerepelnek az adatok. Természetesen előhívható a költségeltérésre vonatkozó jelentés elemző, kiegészítő magyarázata is.

### 7.2.2. A vezetői munka követése

A Commander lehetővé teszi a vezető munkáját támogató automatikus nyomkövetést. Erre a célra a *Reminder* nevű komponens szolgál. A Reminder valójában egy naptár, ami a dokumentumok mellett dátum szerint tárol postázási információkat. Automatikusan emlékezteti a felhasználót, ha az adott napon lejár valamelyik tétel határideje.

A Reminder előnyösen használható költségirányzatok és más tervek készítésekor. Amikor a vezetőség elfogad egy tervet, akkor megjelöli a tervben szereplő fontosabb határidőket. A munkaállomáson látható fontosabb képernyő-képeket el lehet küldeni a Reminder-nek, hogy az a megjelölt határnapon, vagy egy munkaértekezlet időpontjában jelenítse meg a képernyő-képeket. A Reminder figyelmeztet arra, hogy a megadott határnapokon a terv aktuális állapota megismerésre kerüljön, bekérjék a tervadatokat, azaz a tervet naprakész állapotba hozzák.

A Commander Reminder javítja a vezetés eredményességét, gondoskodik az elektronikus posta és a munkaállomások egyéb szolgáltatásainak hatékony igénybevételeéről.

### 7.2.3. Dinamikus vizsgálat

A dinamikus adatvizsgálatok elvégzéséhez a Comander *Execu-View* nevű komponense ad segítséget.

Ha egy vezető szembetűnő eltéréseket tapasztal egy jelentésben, akkor arra törekszik, hogy felderítse annak okát. A valódi helyzet pontos megismerése érdekében igényli, hogy az adatokat különböző perspektívákból szemlélhesse. Nem elég annak ismerete, hogy a vállalatnál a teljes nyereség 5%-kal elmaradt a tervezettől. Lehetnek területek, amelyeken a tervezettnél nagyobb a nyereség. Ekkor viszont lehet olyan terület is, ahol az 5%-nál jóval nagyobb elmaradás létezik. Ezt meg kell vizsgálni, további információkra van szükség. Olyan kérdésekre kellhet keresni a választ, mint: Mi az elmaradás oka? Esetleg egy rosszul működő elosztási csatorna? Versenyképes-e a vizsgált termék? Okozhatja-e az elmaradást egy földrajzi terület vásárlási visszaesése? Vagy egy gyártósor hibája? Esetleg egy fogyasztói csoport megváltozott magatartása?

A Commander többszintű vizsgálatokat támogat, kezdve a Briefing Book egyszerű, előre kijelölt vizsgálataitól a nagyobb mélységre terjedő

kötetlen vizsgálatokig, melyek igen bonyolultak, és írásos jelentések használata esetén szinte kivitelezhetetlenek lennének.

Sok vezető szeretné minél előbb megtudni, hogy egy termék sikeresnek bizonyult-e vagy sem. A folyamatosan érkező feldolgozatlan pénzügyi adatok azonban nem adnak ehhez valódi támpontot. Ami a naplófőkönyvet illeti, az bőségesen tartalmaz adatot, de mindez csak egy könyvelési eszköz, ami nem a vezetőség kiszolgálására készült. A pénzügyi nyilvántartás a nyereséget az üzletágak szerint tartja nyilván, nem pedig a termékek, ill. projektek szerint. Ugyanakkor a mélyebb vezetői vizsgálódás azt követeli meg, hogy a nyereségességet más stratégiai perspektívákból nézve számítsák ki. Ehhez egy többdimenziós adatbázisra van szükség, ami úgy van tervezve, hogy támogassa a pénzügyi teljesítmény vezetői nézetből történő elemzését.

Erre szolgál a Commander *Executive Information Base* nevű adatbázisa. Az ebben rendezett adatok pénzügyi elemzők által meghatározott elvek alapján, automatizáltan kerülnek ide. Például, egy termék nyereségmodellje számára az eladási kilátásokat és a költségkilátásokat az üzleti egység főkönyvéből kell megállapítani, ahol a vállalat termelési költségei a termékek között szét vannak osztva. Eredményül olyan pénzügyi adatokhoz jutunk, amelyek stratégiai kilátásokat jelentenek.

Ha már kialakították a szükséges információs modellt, a vezetők és a pénzügyi elemzők az *Execu-View* felhasználásával könnyű és hatékony hozzáférést nyernek az Executive Information Base-hez. Az Execu-View grafikus interfészét igen egyszerűen, intuitív módon tudják kezelni, egér segítségével.

Az Execu-View eléri a központi számítógépen tárolt Executive Information Base-t és a PC-ken levő helyi adatbázisokat is. Ezekben megtalálható az összes aktív adatváltozat, beleértve a költségelőirányzatot, az aktuális adatokat, a múlt adatait, valamint az összetett előrejelzéseket. Mindezek alapján különböző felületeken, különböző metszetekben lehet vizsgálni a vállalat működését. Lehetőség nyílik a többdimenziós OLAP-műveletek kihasználására, a „szeletelésre” és „forgatásra”, és ezáltal előzetes tervezés nélkül lehet különböző kérdésekre választ kapni, sejtések alapján lehet adatokat követni, rejtett összefüggésekre lehet rájönni. Különösen alkalmas ez az eszköz az elemző habitusú vezetők számára.

Az Execu-View segítségével például végig lehet tekinteni minden főosztály tényleges és tervezett költségeit. Ahol kirívó a költségtúllépés, ott további vizsgálódásra, leásásra és OLAP-elemzésre lesz szükség. Legyen a

példánk egy gyártó vállalat elemzése, ami egy évet fog át. Az adatokat a 7.1. táblázatban láthatjuk.

**7.1. táblázat:** A vállalat 4-ik negyedéves termelési adatai

Termék: Összes		Változó: Bevétel		Negyedév: 04	
Részleg	Költség	Terv	Eltérés	Százalék	
1. részleg	2083	2041	-42	-2	
2. részleg	214	218	4	2	
3. részleg	810	762	-48	-6	
4. részleg	2089	2323	234	11	
<b>Összes részleg</b>	<b>5196</b>	<b>5344</b>	<b>148</b>	<b>3</b>	

Itt látható, hogy a 3. részlegnél van a legnagyobb gond. Ilyenkor le lehet hívni egy táblázatot, amely tartalmazza a 3. részleg költségfelhasználását, termék szerinti bontásban. Ez a 7.2. táblázat.

**7.2. táblázat:** A 3. részleg termelési adatai

Részleg: 3. részleg		Változó: Bevétel		Negyedév: 04	
Termék	Költség	Terv	Eltérés	Százalék	
A termék	189	185	-4	-2	
B termék	150	151	1	1	
C termék	352	304	-48	-14	
D termék	119	122	3	3	
<b>Összes termék</b>	<b>810</b>	<b>762</b>	<b>-48</b>	<b>-6</b>	

A 7.2. táblázatból kiderül, hogy a C termék költségei lépik leginkább túl a tervet. Most meg kell tudni, hogy a túllépés az aktuális negyedévre, vagy az egész évre jellemző. A következő lépés megtekinteni a C termék terv- és tényadatai közötti eltérést, negyedév szerinti bontásban. Ez lesz a 7.3. táblázat. Ebben a táblázatban egyedül a C terméket vizsgáljuk, ahol az adatok a tervszámtól való eltérés százalékában vannak kifejezve.

**7.3. táblázat:** A 3. részleg éves termelési adatai egy termékre**Termék: C termék      Részleg: 3. részleg      Verzió: %-eltérés**

Változó	Negyedév:			
	01	02	03	04
Eladott egységek	-6	8	-12	-20
Bevétel	-5	5	-10	-14
Termékköltség	1	3	8	5
Nettó nyereség	-3	1	-12	-20

A negyedév szerinti bontásban megjelenített adatokból kitűnik, hogy a C termékkel kapcsolatos problémák nem korlátozódnak az aktuális negyedévre. A bevétel az egész évben kisebb a tervezettnél, sőt a tiszta nyereség sokkal jobban elmarad a tervezettől, mint a bevétel.

Ezzel a néhány vizsgálattal meg lehetett ismerni a cég egy részlegével és egy termékével kapcsolatos problémáját, sőt érzékelni lehetett a probléma súlyát is. Ezzel a szabad formájú folyamattal nyert információ rutin-szerűen ad kiindulást a kérdések megválaszolására, továbbá a problémák azonosítására és megoldására.

Az Execu-View a felhasználó számára nem csak táblázatos, hanem sokoldalú grafikus megjelenítést is biztosít, egér és billentyűzet használatával. Ebben grafikonok, oszlopdigramok, tortadiagramok egyaránt kérhetőek. Mindez pedig színes kivitelben.

Az Execu-View közvetlen kapcsolatot teremt a Commander és a cég meglévő SQL-adatbázisai között. Ez a szolgáltatás a vezetőknek numerikus adatok esetén strukturált, többdimenziós lekérdezési és elemzési lehetőséget biztosít. Ez különösen előnyös azoknak, akik hasznosítani akarják az SQL-adatbázis közvetlen elérési lehetőségét, de nem igénylik a modellezési lehetőséget. Tipikus alkalmazás erre az eladások elemzése és a költségek lekötése.

A dinamikus vizsgálat a VIR legfontosabb része. Az üzleti folyamatok aktív vezetői ismerete jó hatással van az egész vállalat nyereségére. Ha a vezetés jobban felhasználja a stratégiailag fontos információkat, akkor minőségileg jobb döntéseket hoz. A VIR segít a vállalat üzleti információi közötti kapcsolatok megértésében, ezáltal növeli a nyereséget és javítja a működést.

#### 7.2.4. Kommunikáció másokkal

Egy VIR használata során gyakran merülhet fel olyan igény, hogy egy vezető meg szeretné osztani gondolatait, vagy felvetéseit, problémáit másokkal. Ezekre nyilván választ is vár. Itt szóba jöhet például egy képernyőkép, amit szeretne elküldeni egy vagy több kollégájának, akik nincsenek elérhető közelségben. A Commander erre az igényre is rendelkezik egy komponenssel, amit *Redi-Mail*-nek neveztek el.

A Redi-Mail egyesíti a VIR-működést az elektronikus postázással. Lehetővé teszi, hogy a felhasználó kijelöljön egy képernyőképet a Briefing Book-ban, vagy az Execu-View-ban, vagy bármelyik más komponensben, és azt elláthatja a saját kiegészítő információjával, majd elküldheti azokra a számítógép-címekre, amelyek felhasználóit be akarja vonni a probléma megoldásába.

A Redi-Mail-lel továbbítható egy értekezletről készült feljegyzés, vagy a feladó egyszerűen közölheti gondolatait a címzettekkel. Mindezt a Commander által adott megjelenési formában, eltérően az E-mail rendszerek általános formájával.

A Redi-Mail Commander-stílusú elérést és feldolgozást biztosít a beérkező és az archivált dokumentumokhoz.

#### 7.2.5. Kapcsolat a külső világgal

A vezetésnek fontos feladata, hogy előre meglássa, mikor kell változtatni a szervezet taktikájában és stratégiájában, annak érdekében, hogy cég megőrizze versenyképességét. A változtatásokat elindító legkorábbi jelzések nem belülről, a belső adatokból származnak, hanem abból az információból, ami a többi versenytárs tevékenységének változására, a beszállítók helyzetére és terveire, valamint a piaci környezet változására vonatkoznak. Olyan külső változásokra, amelyek lényeges hatással lehetnek a saját vállalat teljesítményére.

A leírtakkal összhangban, a Commander olyan komponenssel is rendelkezik, ami a vállalaton kívüli világgal való információs kapcsolatok tartására szolgál. Ez a *News Navigator*. A News Navigator elősegíti a vállalat környezeti változásainak követését. Néhány olyan terület, aminek a figyelését lehetővé teszi:

- A versenytárs teljes sajtóanyaga. Mindent el kell olvasni, amit a cégek magukról közölnek, nem csak azt, amit az újságok és magazinok írnak.

- A kiugró fontosságú új hírek, amelyek kapcsolatban állnak az iparág változásaival, trendjeivel, a világ eseményeivel, és több más olyan téma, amely befolyásolhatja a vállalt terveit és céljait.
- Az iparággal vagy más cégekkel kapcsolatos működési adatok, összehasonlító értékelések.
- Cégek tőzsdei értékelésére, a részvényeire vonatkozó információ, beleértve a fontosabb tőzsdeindexeket. (Például, az USA-ban a Dow Jones vagy a Nasdaq.)
- Olyan nyilvános adatok, amelyek érdekesek lehetnek a Commander navigációs műveletei során. Lehetőség van arra, hogy több mint 450 publikációból tudjunk híreket keresni. A News Navigator ehhez számos segédeszközt tartalmaz, amelyek a keresést könnyítik és gyorsítják.
- Mód van arra is, hogy kiválasztott cégekről külön dossziét tartsunk fel, amelyben a befektetők összetételét, a részvények megoszlását, mérleg-eredményeket, és más kulcsfontosságú adatokat tároljunk.

A Commander News Navigator-ával el lehet érni az összes lehetséges híryanagot, amit az üzleti világ számára rendelkezésre álló források közölnek elektronikus formában. Ennek megvalósítása a számítógépes dokumentumkezelési folyamatok végrehajtását is megköveteli.

A News Navigator össze van építve a többi Commander alkalmazással. Ennek megfelelően a felhasználó jegyzetekkel láthatja el a begyűjtött információt, és az egészet elmentheti a Reminder-ben, későbbi megjelenítés céljából, vagy pedig elküldheti azt a Redi-Mail-lel, hogy felhívja mások figyelmét is az érdekesebb adatokra.

Az üzleti információk felhasználhatók arra, hogy a számítógépen vállalatokat lehessen összehasonlítani egymással, vagy iparági és piaci átlagadatokat összevetni. Mindehhez a Commander összes rendelkezésre álló grafikus eszköze is segítséget ad.

### 7.2.6. OLAP elvégzése

A Comshare szoftverház *Commander OLAP* nevű terméke egy későbbi fejlesztési fázis eredményeként került be a VIR-komponensek közé. Ez a komponens az OLAP célú feldolgozást teszi lehetővé.

Az OLAP komponens egy olyan szoftvercsomag, amely egy sokdimenziós modellezőt tartalmaz, beépített SQL-interfészsel és külön pénzügyi modullal.



Ezen kívül még a következő eszközöket tartalmazza:

- Az információ megjelenítéséhez szükséges modul.
- Hálózati felügyelő szoftvert.
- Adatszűrő és adatintegráló modul.

Ezeknek az összetevőknek az integrálása lehetővé teszi, hogy a vállalati vezetés egy sokkal szélesebb rétegét szolgálja ki a Commander OLAP, annak köszönhetően, hogy a vállalati információs rendszer szinte mindegyik szintjéhez közvetlen hozzáférést, azonnali (*on-line*) analízist tesz lehetővé.

Az OLAP-ban felhasznált adatbázis-kezelő az *Essbase*, ami a közvetlen, sokdimenziós analitikai feldolgozás számára készült, ugyanakkor pedig egy vállalat különböző szervezeti egységeinek a VIR-feladataihoz *on-line* áll rendelkezésre.

Az *Essbase* adatbázis kliens-szerver módban biztosítja az azonnali hozzáférést és a sokdimenziós analízist. Az adatok származhatnak szinte bármilyen forrásból: relációs adatbázisokból, táblázatokból, vagy akár egyszerű ASCII fájllokból. A döntéshozók az *Essbase*-zel az adatokat tetszőleges részletességgel és bármilyen szempont szerint rendezve, intuitíven kezelhető környezetben tekinthetik át.

Üzleti siker csak akkor érhető el, ha figyelembe vesszük az összes tényezőt, mint például: termékek, területek, elosztási csatornák, mennyiség stb. Mint tudjuk, a sokdimenziós analízis azt jelenti, hogy lehetőség van megtekinteni egyidejűleg minden olyan változót, ami befolyásolhatja a döntést. Az *Essbase*-zel az adatok közötti navigáció nagyon könnyen megvalósítható, és lehetőség van arra is, hogy különböző *ad-hoc* számításokat végezzünk (például trendeket), amik fontosak lehetnek a döntésekhez.

Az adatok különböző adattárolókból, táblázatokból, relációs adatbázisokból, számviteli rendszerekből, egyszerű egérmozgatással kerülnek az *Essbase* modellbe. A felhasználók időt takarítanak meg az adatgyűjtésnél, integrálásnál és konszolidálásnál, így több idejük marad az analízisre.

Egy vállalat működéséhez szükséges adatok nagy teljesítményű rendszert igényelnek. Az *Essbase* könnyen kezeli a nagyobb adattömeget is (milliárdos nagyságrendben), különböző tömörítő módszereket alkalmazva, és maximálisan kihasználva a tárolókapacitást, aminek az eredménye a hozzáférés a teljes információhoz, ami szükséges az alapos elemzéshez. Az *Essbase* egyesíti a legkorszerűbb felhasználói, hálózati és adatbázis szintű eszközöket. Azok a felhasználók, akik más hálózatban vannak, egyszerűen elérhetik az *Essbase* adatbázist TCP/IP vagy egyéb ismert protokollal.

Fontos még megjegyezni, hogy az Essbase olyan termék, amely megfelel mind a tizenkét Codd-féle OLAP-szabálynak.

### 7.2.7. Szakértői rendszer alkalmazása

Az 1.7. fejezetben már áttekintettük a szakértői rendszerek szerepét és jelentőségét a szoftver eszközök körében, ezen belül a vezetői információs rendszerek vonatkozásában is.

A Comshare Corporation törekedett arra, hogy a Commander terméket is ellássa ilyen segédeszközzel. Az erre a célra létrehozott szakértői rendszer elnevezése: *Detect & Alert*.

A Detect & Alert arra szolgál, hogy a felhasználó által külön kijelölt folyamatokat állandóan kövesse, és jelzést adjon abban az esetben, ha a folyamatban valamilyen nem kívánatos esemény, változás következett be.

A megfigyelés kiterjedhet kritikus adatforrások tartalmára, olyan dolgok bekövetkeztére, amiket a felhasználó tudni akar, és amikről jelzést vagy figyelmeztetést kíván kapni. Ezt a funkciót ún. *szoftver robotok* látják el. Az automatizált megfigyelés azzal jár, hogy minden egyes lehetséges szituáció ellenőrizve lesz, mielőtt valamilyen változás lép fel a vele kapcsolatos adatokban.

A szakértői rendszer összegyűjti a robotok által küldött jelzéseket, azokat prioritási sorrendbe állítja, és közli a felhasználóval. A tájékoztatás szöveges formában történik, a felhasználó számítógépének képernyőjén.

A szoftver által végzett megfigyelés az alábbiakra terjed ki:

- A kritikus sikertényezőkre (KST).
- A kulcsfontosságú teljesítménymutatókra (KTM).
- A nyereségességre, árakra, árengedményekre.
- Bevételi előrejelzésekre.
- Az ügyfelek elégedettségére.
- Minőségi elemzésekre.
- Termelési előrejelzésekre.
- Költségvetési elemzésekre és jelentésekre.
- Raktárkészletekre.
- Termék, üzletág, régió szerinti nyereségességre.
- Termelési ág nyereségességre.
- Eladási csatorna elemzésére.
- Költségvetési jelentésekre.
- Versenytárs figyelésére.
- Piaci eladások elemzésére.

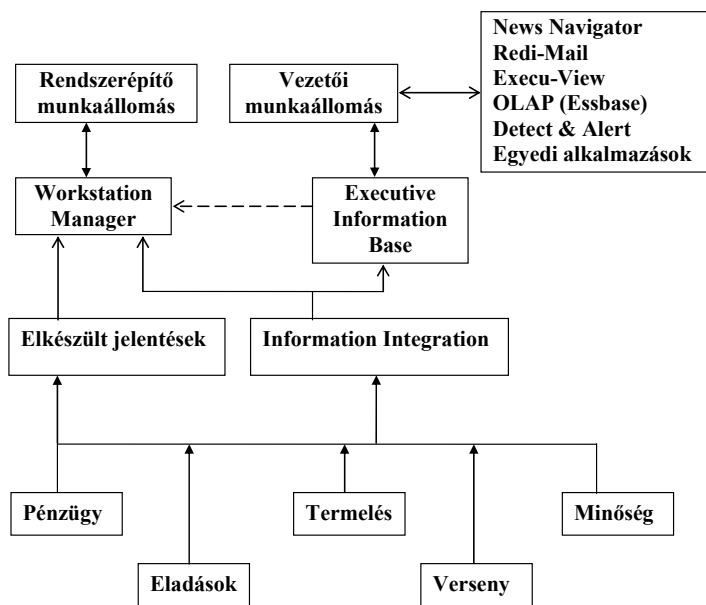
### 7.2.8. A Commander információs folyamatai

A Commander üzemelése hálózati kapcsolatban álló számítógépeken valósul meg, ahol az egyes szoftverfunkciók elosztása a működtetési kényelem és hatékonyság célját szolgálja.

A Commander szoftvere egyrészt a nagyteljesítményű központi számítógépen, másrészt pedig az ehhez kapcsolódó munkaállomásokon, PC-ken üzemel. Azok a funkciók, amelyek hatékonyabban végezhetőek el a központi gépen, azon hajtódnak végre, míg a többi funkció a PC-kre van kihelyezve.

A központi gépen működik a Commandernek a teljes működést irányító komponense, az *Information Integration*, a munkaállomások tevékenységét szervező *Workstation Manager*, valamint itt helyezkedik el az *Executive Information Base* nevű központi adatbázis is. Az előző pontokban leírt Commander-komponensek mindegyike a központi gépen funkcionál. Ugyanakkor azonban az egész rendszer úgy van megvalósítva, hogy a központi gép és a munkaállomások közötti együttműködési kapcsolat egyes funkcióknak a PC-kre, történő kihelyezését is lehetővé teszi, amikor az lehetséges, attól függően, hogy melyik megoldás a hatékonyabb a felhasználó vezetőik szempontjából.

A Commander működésének vázlatos áttekintésére szolgál a 7.1. ábra.



7.1. ábra: A Commander felépítése és kapcsolatai

Az ábrán a felhasználói és egyéb munkaállomásokat egyetlen kiragadott állomással képviseltettük. A *rendszerépítő munkaállomás* szoftver-eszközei lehetővé teszik, hogy a vezetői információs modellt, az alkalmazási prototípust ki lehessen dolgozni, és meg lehessen valósítani. Ezeket az eszközöket a vállalat adatszolgáltatói és a VIR-t testre szabó fejlesztők egyaránt hatékonyan tudják használni.

Megjegyzések:

A Commander rendszer első generációjának elterjedésével és felhasználásával kapcsolatosan még az alábbi adatokat ismertetjük:

- A Gartner Group (USA) 1993-as átfogó felmérése szerint több mint 500 teljesen kifizetett Commander-állomás működött a világon. A felhasználók száma ezeknél meghaladta a 20 ezret, és ez a szám ez után is gyorsan növekedett.
- A központi számítógép elsősorban az IBM AS-400-as rendszere volt, másodsorban pedig a DEC VAX gépei látták el a teljesítményigényes funkciókat.

### 7.3. Néhány jellemző Comshare VIR-alkalmazás

Az alábbiakban rövid kifejtéssel áttekintünk néhány olyan Comshare-alkalmazást, ami jól tükrözi a szóban forgó szoftver fontosabb használati területeit.

**Bankok:** Szabályzati és irányítási jelentéseket kezelő alkalmazások, testületi, pénzügyi és ügyfélszolgálati információk a közép- és felsőszintű vezetők számára.

**Légitársaságok:** Minőségellenőrzési jelentéseket generáló alkalmazások, beleértve az érkezési időket, utas-kiszolgálást, és az egyes járatok nyereségességi mutatóit.

**Vízművek:** Vezetői információs rendszer segítségével követik nyomon a vízminőség-változásokat, valamint a vízellátást, továbbá a fogyasztói és szerviz-információkat.

**Ipari minőségellenőrzés:** Korszerű on-line monitor-rendszerrel követhető a gyártástechnológia tetszőleges fázisa.

**Biztosítók:** A világ három legnagyobb biztosítási vállalkozása használ Comshare rendszereket, az egész világot átfogó üzleti tevékenységük konszolidációs kimutatásához.

**Kereskedelem:** Több kereskedelmi lánc használ Comshare VIR-t a napi eladások (termékek és elárúsítók helyek szerinti) áttekintésére, kiértékelésére, ill. vásárlói igénynek előrejelzésére.

## 7.4. A Comshare második generációs termékei

### 7.4.1. Fejlesztési irányok

A Comshare Corporation folyamatosan fejlesztette tovább a VIR-területhez tartozó termékeit. Az 1995-től elérhető újabb változatok is folyamatos bővítésen, újabb verziók kibocsátásán mennek keresztül. Ezek a termékek önállóan is alkalmazhatóak, de az igényektől függő együttes, összekapcsolt felhasználásuk is lehetséges.

Az újabb generáció szoftver eszközei már erősen igazodnak a modern számítógép-hálózatok világához, a kliens-szerver felépítéshez, valamint az Interneten alapuló Web-fejlesztésekhez, Web-környezetekhez.

A Comshare termékek háromszintű kliens-szerver struktúrájú megoldást igényelnek. Ennek ajánlott konfigurációja a következő:

*Szerver:*

- Pentium PC, DEC Alpha, IBM RISC, SUN
- Microsoft Windows, AIX, HP és SUN Unix

*Adatbázis-kezelés:*

- OLAP multidimenzionális (Essbase, Oracle Express, TM1, Microsoft Plato)
- Relációs (Oracle, Microsoft SQL Server)

*Kliens állomás:*

- Középkategóriájú Pentium PC
- Microsoft Windows

*Hálózat:*

- TCP/IP

A cég nemcsak az USA-ban, hanem az egész világon forgalmazza termékeit. Jelenleg több mint 3000 vállalat tartozik az ügyfélkörébe. Ezek között számos világcég is megtalálható.

Magyarországon az egyedüli képviselőt az *SZKI Comshare Kft.* látja el. A VIR-ek kialakítását is ez a vállalkozás végzi. Az első hazai alkalmazás a BKV-nál valósult meg 1996-ban. A BKV VIR a közlekedési vállalt teljes működési keresztmetszetére kiterjed. Ez az alkalmazás a *Comshare Decision* nevű második generációs terméken alapszik.

A legnagyobb méretű kiépítés a MOL Rt.-nél valósult meg. Ez mintegy 100 felhasználót foglal magában, a MOL-nak az SAP informatikai rendszeréhez kapcsolódóan. Az itteni VIR-ben teljesítményellenőrzési, beszámolási és elemzési alkalmazások működnek.

A következőkben ismertetjük az egyes Comshare termékek főbb működési jellemzőit.

#### **7.4.2. A Comshare Decision**

A *Comshare Decision* a kliens-szerver üzemelést és a Web-technológiát megvalósító, nyitott architektúrával rendelkező döntéstámogató rendszer. Felhasználja a legelterjedtebb adatbázis-kezelőket, és adatáruház megoldásokat, korszerű adatmegjelenítést alkalmaz, különböző típusú közgazdasági elemzési lehetőségeket nyújt.

A Decision a számítógépes üzleti intelligencia eszközeit felvonultató keretrendszer, amely az aktuálisan szükséges tartalommal feltöltve, igény szerinti VIR modulokat valósít meg. Ilyen modulok lehetnek például azok, amelyek a

- teljesítményirányítás,
- beszámolás,
- elemzés,
- benchmarking,
- stratégiai tervezés,
- humán erőforrás-gazdálkodás,
- befektetés stb.

területén kerülnek felhasználásra.

A Comshare Decision hatékony, optimum-orientált, grafikus megjelenítő, elemző és modellező eszköz, ami igen gyors adatelérést biztosít az *ad-hoc* elemzésekhez, összefüggések feltáráshoz, döntési alternatívák és következmények gyors kiértékeléséhez. Lehetőséget ad arra, hogy egy kialakított üzleti modell adatait vezetői, tervezői, ill. elemzői igények szerint vizsgálhassuk és megjeleníthessük. Mindezt többdimenziós adatbázis-kezelés révén valósítja meg.

A rendszernek létezik olyan változata, amely Web-alapú környezetben is működik. Ez a Comshare DecisionWeb, ami a szokásos Web-keresőkhöz (browser-ekhez) hasonló keretben üzemelhet.

### **7.4.3. A Comshare BudgetPlus**

A *Comshare BudgetPlus* a vállalati költségvetés tervezési folyamatának elvégzésére szolgál. Biztosítja a költségterv integritását, és a vállalat konszolidált működését segíti elő.

A rendszer a tervezés, elemzés és ellenőrzés támogatásán túlmenően biztosítja a különböző alternatívák, trendek és összefüggések feltárását, jelzi a kedvezőtlen feltételeket. A tervezés során az egyes részterületek terveinek koordinálása és konszolidálása révén a BudgetPlus használatával a tervezési idő és ráfordítás jelentős csökkentése érhető el, továbbá a különböző tervváltozatokból a legjobb terv kiválasztása történhet meg.

A BudgetPlus a Decision technikai eszközrendszerére épül. A két rendszer integrált használata nemcsak lehetséges, de egyenesen célszerű és ajánlatos.

### **7.4.4. A Comshare FDC**

A *Comshare FDC* a vállalat LAN-ján működik, a pénzügyi számítások elvégzésének és az ezekhez kapcsolódó vezetői jelentések elkészítésének hatékony eszköze. Magában foglalja a táblázatkezelő szoftverek összes funkcióját.

A különböző formátumú számlák alapján is képes összesítéseket készíteni, összeköthető a naplófőkönyv adatkészletével. Grafikus megjelenítéssel lehet pénzügyi elemzések, összehasonlítások eredményeit közölni.

A Comshare FDC integrálható a BudgetPlus szoftverrel. A két eszköz együttműködése közös adatbázis felhasználásával történik.

## Irodalomjegyzék

- Dr. Abonyi János (szerk.): *Adatbányászat, a hatékonyság eszköze*, ComputerBooks, Budapest, 2006.
- P.Adriaans, D. Zantinge: *Adatbányászat*, Panem Könyvkiadó, Budapest, 2002.
- Michael J. A. Berry, Gordon S. Linoff: *Mastering Data Mining, The Art and Science of Customer Relationship Management*, John Wiley & Sons, Inc., USA, 2000.
- Earl Cox: *Fuzzy Modeling and Genetic Algorithms for Data Mining and Exploration*, Elsevier Inc., Morgan Kaufman Publishers, USA, 2005.
- Executive Information Systems*, Management Guides, Business International Limited, United Kingdom, July, 1991.
- J. Han, M. Kamber: *Adatbányászat*, Panem Könyvkiadó, Budapest, 2004.
- Heckenast Tamás: *Üzleti információ vizualizálása*, in *Controllingtrendek*, Raabe Tanácsadó és Kiadó Kft., 2006, 5/2/1-26. old.
- Hetyei József (szerk.): *ERP rendszerek Magyarországon a 21. században*, ComputerBooks, Budapest, 2004.
- Hetyei József (szerk.): *Vállalatirányítási információs rendszerek Magyarországon*, ComputerBooks, Budapest, 1999.
- Hetyei József (szerk.): *Vezetői döntéstámogató és elektronikus üzleti megoldások Magyarországon*, ComputerBooks, Budapest, 2001.
- Hetyei József: *Vezetői döntéstámogató és elektronikus üzleti megoldások Magyarországon*, 2001.
- Henry F. Korth, Abraham Silberschatz: *Database System Concepts*, McGraw-Hill, Inc., Singapore, 1991.
- Michael Negnevitsky: *Artificial Intelligence: A Guide to Intelligent Systems*, Addison-Wesley Publishing Company, Great Britain, 2002.
- James A. O'Brien: *Introduction to Information Systems: An Internetworked Enterprise Perspective*. Irwin/McGraw-Hill 1998.
- John F. Rockart, David W. DeLong: *Executive Support Systems: The Emergence of Top Management Computer Use*, Business One Irwin Company, USA, 1988.
- Robert Schultheis, Mary Sumner: *Management Information Systems: The Manager's View*, McGraw-Hill Company, USA, 1998.
- Ian Sommerville: *Software Engineering*, Seventh Edition, Pearson Education Limited, USA, 2004.



- Ian Sommerville: *Szoftver-rendszerek fejlesztése*, Software Engineering, (Sixth Edition), PANEM Könyvkiadó, Budapest, 2002.
- Dr. Sziray József, Gaul Géza, Erdős Ferenc: *Vállalati információs rendszerek II.*, Egyetemi jegyzet, Universitas Kht, Győr, 2006.
- Dr. Sziray József, Gaul Géza: *Vállalati információs rendszerek I.*, Egyetemi jegyzet, Universitas Kht, Győr, 2006.
- Dr. Sziray József: *Szakértői rendszerek alapjai*, Egyetemi jegyzet, Universitas Kht, Győr, 2006.
- Dr. Sziray József: *Vállalati információs rendszerek*, KHVM-tanulmány, Széchenyi István Egyetem, Győr, 2002.
- Zhengxin Chen: *Computational Intelligence for Decision Support*, CRC Press LLC, USA, 2000.

## Rövidítések

A könyvben előforduló rövidítések (cég- és terméknevek nélkül):

Rövidítés	Kifejtés	
2D		kettő dimenziós geometriai ter- rek
3D		három dimenziós geometriai terek
BPR	business process reengi- neering	üzleti folyamatok újjászervezése
CRISP-DM	Cross Industry Standard Process for Data Mining	iparágaktól független adatbányá- szási módszertani szabvány
CSF	critical success factor	kritikus sikertényező (L. még KST!)
DBMS	data-base management system	adatbázis-kezelő rendszer
EIS	executive information sys- tem	vezetői információs rendszer (VIR)
FASMI	fast analysis of shared multidimensional infor- mation	megosztott többdimenziós in- formáció gyors elemzése
GDP	gross domestic product	bruttó hazai (nemzeti) összter- mék
HCI	human computer interac- tion	ember-gép kapcsolat (eredetileg kölcsonhatás)
HOLAP	hybrid OLAP systems	vegyes OLAP (L. ott!) rendsze- rek
IT	information technology	információ-technológia <i> vagy </i> infor- máció-technika
KDD	knowledge discovery in databases	adatbázisokban végzett tudásfel- tárás
KPI	key performance indica- tor	kulcsfontosságú teljesítmény mutató (L. még KTM!)
KST	kritikus sikertényező	angolul critical success factor (CSF)
KTM	kulcsfontosságú teljesít- ménymutató	L. még KPI!
MIS	management information system	menedzsment információs rend- szer

MOLAP	multidimensional OLAP systems	multidimenzionális OLAP (L. ott!) rendszerek
NASDAQ	National Association of Securities Dealers Automated Quotations system	elektronikus tőzsde rendszer
OLAP	on-line analytical processing	azonnali elemzés, többdimenziós adatbázist használó feldolgozási mód
OLTP	on-line transaction processing	üzleti tranzakciók azonnali gépi feldolgozása
RI	rule interest	kovariancia analízisnél használatos mutató
ROI	return of investment	tőkehozam
ROLAP		relációs adatbáziskezelőn alapuló OLAP
SW	software	szoftver
SWOT	strengths, weaknesses, opportunities, threats	erősségek, gyengeségek, lehetőségek, veszélyek (GYELV), stratégiai elemzési módszertan
SzR	szakértői rendszer	
TID	transaction identifier	ügylet (tranzakció) azonosító
URL	Uniform Resource Locator	egységes alakú forráshely megadás (Interneten használt)
VIR	vezetői információs rendszer	L. még EIS!
ZVEI	Zentralverband der Elektroingenieure	Német Villamosmérnökök Központi Egyesülete