

A STATISZTIKAI ADATBANKOK TERVEZÉSÉNEK EGYES ELMÉLETI KÉRDÉSEI

ESZTERGÁR ZSOLT – WILLIAM F. SMYTH

Napjainkban világszerte fokozódó érdeklődés mutatkozik az információfeldolgozás területén az adatbank vagy adatbázis-kezelő rendszerek irányában. Nagy információszervező rendszerek kidolgozása nehezen valósítható meg ezen rendszerek nélkül.

Bár a téma ma már túl van azon, hogy kizárólag elméleti kutatások tárgya legyen – sok helyen működnek hatékony adatbázis-kezelő rendszerek –, mégis érdemes lehet néhány elméleti probléma felvetése és vizsgálata egy speciális témakör, a statisztikai adatbázis-kezelő rendszerek vonatkozásában.

A feladat több irányból is megközelíthető, például a statisztikai információrendszerek kialakításának munkája során megoldandó feladatok vagy egy hatékony feldolgozási rendszer kidolgozása érdekében megoldandó problémák szempontjából.

A következőkben a statisztikai adatbankok tervezése során – elsősorban a feldolgozási rendszer hatékonysága szempontjából – felvetődő néhány problémát vizsgálunk meg.

AZ ADATOK SZERKEZETE

Adaton olyan a elemek (elemi információk) halmazát érthetjük, melyeket egy d leírás és egy q előjeles érték egyértelműen specifikál. Így:

$$a = (d, q)$$

A d értelmezése lehet például az 1969-ben Magyarországról Jugoszláviába exportált kerékpárok száma és a hozzá tartozó $q = 9153$.

Általában, a tároló gazdaságosabb kihasználása érdekében, d helyett egy c kódot szokás használni. Így

$$a = (c, q)$$

A fenti példa szerint így a következőt kaphatjuk:

$$a_j = (c, q_j)$$

$$a_j = (1497209410669, 9153),$$

ahol c_j az 1. táblában leírt módon értelmezhető.