

a „hard copy” (papírmásolat) output.

Az adatbázis nagysága problémát okoz, ha a rendszerben nincs olyan nagykapacitású tároló egység, amelyben az egész file elférne, ha bizonyos file-okat nem lehet invertálni, ha az indexelés inkonzisztens, ha a formátumok sőt a kódolás módja változott az évek során, ha nagyon sok változó írja le a komplex irodalomkutatás témáját, ha a file hierarchikus szervezésű stb.

#### Gazdaságosság

Becslések szerint az on-line működtetéshez szükséges inverz tárolás kb. százszor költségesebb, mint a mágnesszalagos. A nagyméretű adatbázisokkal kapcsolatos minden más manipuláció (előállítás, újraszervezés, invertálás stb.) is rendkívül költséges. Csak akkor válik gazdaságossá az adatbázis működtetése, ha különböző szolgáltatások egyaránt felhasználhatják azt. Ez nagyfokú normalizálást igényel. Új lehetőségeket rejt magában az öntanuló rendszerek alkalmazása: az ilyen rendszer az igények folyamatos figyelembevételével újraszervezi a file-okat, ezzel csökkentve a soron következő igények kielégítésének költségeit.

#### Kutatás

A nagyméretű integrált adatbázisok nemcsak hatékonyság és gazdaságosság szempontjából előnyösek, hanem sok olyan melléktermékük van, amely nagy jelentőségű lehet. Így például a nagyméretű adatbázisokon végzett statisztikai vizsgálatok segítséget nyújtanak tudománypolitikai feladatok megoldásához, folyóiratok értékeléséhez, szótárak összeállításához stb.

Más vizsgálatok pedig visszahatnak a file szerkesztés elveire. Az egyik legtöbbet ígérő vizsgálat kimutatta, hogy a *Chemical Titles* és CAIN (Cataloging and Indexing) adatbázisokban a 127 leggyakrabban előforduló címszó az összes előfordulásuk 44%-át tette ki. Ezekhez 8 bites kódokat rendelve 63%-os adat-kompresszió érhető el.

Másik lehetséges megoldást nyújt a tömegtárolók alkalmazása. A konferencián ismertettek egy 10<sup>12</sup> bit nagyságú fotodigitális tároló használó rendszert is.

A műszaki és gazdasági nehézségek leküzdésével kialakított nagyméretű adatbázis rendszerek azonban olyan pszichológiai, szociológiai, politikai és egyéb problémákat is felvetnek, amelyek megoldása további nagy erőfeszítéseket igényel.

*Journal of Chemical Information and Computer Sciences*, 15. k. 1. sz. 1975. p. 3–5.)

(Valkó Péter)

\* \* \*

## Környezetügyi adatrendszerek az Egyesült Államokban

*Az Egyesült Államok Legfőbb Állami Számvevőszéke (General Accounting Office, GAO) a kongresszus megbízásából felülvizsgálta a környezetügyi adatok gyűjtésének és tárolásának helyzetét.*

A vizsgálat kiterjedt az adatok fajtáira, az adatgyűjtés körére, az információtárolás módjaira, az információk hozzáférhetőségére a magán- és a közületi szektor számára, beleértve a kormányhivatalokat is; továbbá az adatgyűjtés koordinálásának módszereire és az adatoknak rendszerek közötti cseréjére.

8 minisztérium irányítása alá tartozó 320 adatközponttól és 10 önálló, független intézménytől érkezett válasz. Valamennyi gyűjt és tárol környezetügyi adatokat, többen az egész országra kiterjedően. Jelentős részük ugyanazon a részterületen, pl. 120 intézmény a levegővel kapcsolatosan gyűjt és tárol adatokat, ezeket a legtöbb rendszerben gépi adathordozókon, pl. lyukkártyákon és mágnesszalagokon tárolják.

Bár a szövetségi kormányknak nincs központi intézménye az adatok cseréjének és gyűjtésének összefogására, 167 intézmény jelezte, hogy rendelkezésre bocsát adatokat és információkat, 94 intézmény pedig kap adatokat más rendszerektől. 98 intézmény tagja olyan hálózatnak, amelyek az adatok cseréjére, illetve a gyűjtőtevékenység koordinálására alakultak. Egy esetben jelezték, hogy az adatok cseréje automatikusan, számítógépek között történik. Egyéb használatos eszközök, módszerek: mágnesszalag, kihelyezett terminál, papírmásolat, közvetlen kapcsolatok.

Az adatcserében információs központok felállításával kívánják előrelépni. Terveznek pl. számítógépes masterindexet a vizekre vonatkozó adatforrásokról.

Más rendszerek adatainak beszerzésében és hasznosításában az alábbi nehézségeket sorolták föl az érintettek:

- az egyes rendszerekben használatos adatelemek és kódok nem kompatibilisek;
- az egyes rendszerekben alkalmazott berendezések és információt őrző nyelvek nem kompatibilisek;
- a kért adatokat gyakran késlekedve szolgáltatják;
- a más rendszerekből származó adatok pontosságát, megbízhatóságát nehéz fölmérni;
- az adatok újrakódolása költséges.

*Information Retrieval and Library Automation*, 10. k. 9. sz. 1975. p. 5–6./

(Fazekas Zsuzsanna)

