

Társadalomtudományi folyóiratcikkek visszakeresése

Az ISI (*Institute for Scientific Information - Tudományos Tájékoztatási Intézet*) bejelentette, hogy on-line hozzáférést biztosít a társadalomtudományi folyóirat-cikkek legnagyobb információs tárához, a *Lockheed Információs Rendszerhez*. A rendszer *SOCIAL SCISEARCH* néven 1974 szeptemberétől működik. Az adattár 1200 társadalomtudományi folyóirat összes cikkére, továbbá 2500 természettudományi folyóirattól válogatott társadalomtudományi tárgyú cikkére terjed ki. Az adatbázis, amelyet 1972 óta gyűjtenek kb. 300 ezer tételt tartalmaz; havi gyarapodása mintegy 7000 tétel. A keresés többféleképpen lehetséges: kulcsszó, szótó, szóösszetétel, szerző és intézmény szerint. Hivatkozások szerint is lehet keresni. A *DIALOG* elnevezésű *Lockheed* program interaktív keresést tesz lehetővé. A Kaliforniában működő számítógéphez a terminálok telefonvonalon kapcsolhatók. Minden kikeresett tételről teljes bibliográfiai leírást kap az igénylő, kiegészítve a cikkekben foglalt hivatkozások teljes jegyzékével.

Journal of Library Automation, 8. k. 1. sz. 1975. p. 75-77./

(Sárdy Péter)

* * *

Konferencia a nagyméretű adatbázisok témakörében (USA)

1971-ben az Egyesült Államok Tudományos Akadémiája (*National Academy of Sciences*) kutatócsoportot hozott létre nagyméretű adatbázisok vizsgálatára. Bár a csoport eredetileg a *Kémiai Információs Bizottság* (*Committee on Chemical Information*) keretében alakult meg, a vizsgálatot később nem kémiai jellegű adatbázisokra is kiterjesztették.

A csoport meghatározása szerint *nagyméretű adatbázis*

nagyszámú rekordot, sokféle adatmezőt vagy adatelemet tartalmaz;

lehet nagyméretű azért mert nagyszámú kikereshető, kiválasztható elem van benne;

végül lehet nagy abban a könnyen érthető értelemben is, hogy nagy tároló kapacitást kíván.

A vizsgálat során csak olyan gépileg tárolt adatbázisokra szorítkoztak, amelyek legalább 100 millió karaktert tartalmaznak. Hamarosan kiderült, hogy a tárolás és

információkeresés alapelvei, fejlődési iránya és problémái nagyon hasonlóak a különböző alkalmazási területeken. Ez tette lehetővé és egyben szükségessé az alábbiakban ismertetett első konferencia megtartását, amelyre az USA Tudományos Akadémiáján került sor 1974. május 22-23. között a *National Research Council* (*Nemzeti Kutatási Tanács*) védnöksége alatt.

Nagy file-ok on-line működési módban

A konferencián tartott előadások közül sok foglalkozott az on-line működésű bibliográfiai információkereső rendszerekkel.

Az on-line szolgáltatás sikerét több tényező magyarázza. Egyrészt nőtt a retrospektív irodalomkutatás számára elérhető adatok mennyisége. Így például a *Chemical Abstracts* alapján, amely 1970-ben még csak két év anyagát tárolta géppel olvasható formában. 1974-ben már öt évre visszamenő irodalomkutatást lehetett elvégezni. Ugyanakkor növekedett az on-line szolgáltatás ismerőinek és felhasználóinak száma, fokozódott a felhasználók bizalma. A megnövekedett igény egyben gazdaságosabbá tette a nagyméretű adatbázisok fenntartását és üzemeltetését. Hasonló irányba hatott a technika fejlődése is, különösen a távközlési kommunikációs hálózatok kiépülése. Az on-line szolgáltatás ma már nem kuriózum, hanem kereskedelmi realitás. Becslések szerint 1974-ben 700 000 on-line irodalomkutatásra került sor, 1975-re számuk várhatóan egy millióra nő.

A nagyméretű adatbázisok problémái

A nagy és kis file-oknál egyaránt alapvető fontosságú a biztonságos adatkarbantartás. A véletlen és szándékos adatkárosodás elleni védekezéshez ismerni kell az adattárakat fenyegető veszélyeket.

Másik alapvető fontosságú probléma az adatok ábrázolásához felhasználható karakterkészlet adekvátsága. A kémiában és a biológiában kb. 1000 különböző karaktert használnak, más diszciplínák is hasonló számú karaktert igényelnek. Az egyes diszciplínák karakterkészletei azonban nem teljesen fedik át egymást. Ugyanakkor kimutatható az is, hogy például a kémiai szövegekben az összes előfordulás 94,5%-ában 36 alapjel szerepel (figyelmetlenül kívül hagyva a jel előfordulásának körülményeit: index, kisbetű, nagybetű stb.). A rendszerben használható karakterkészlet kialakításánál azonban figyelembe kell venni az input/output egységekkel és más file-okkal, szabványokkal való kompatibilitás követelményeit is.

A kémiai információ-tároló és kereső rendszerekben nehezen megoldható feladatot jelent a kémiai vegyületek egyértelmű megnevezése, azonosítása is.

A tapasztalatok szerint az információs rendszer szűk keresztmetszete minden esetben az input adatrögzítés és