

gépes adatállomány-mutatója, amely mintegy 50 USA-beli szövetségi hivatal, intézmény több mint 1300 értékes numerikus és szöveges adatállományát tartalmazza.

#### Mi várható a jövőben?

Az online ipar olyan újdonsága, mint a CD-ROM technika, várhatóan az adatbázisok online forgalmára hat vissza. CD-ROM-on fogják használni a ritkán aktualizálható adatbázisokat.

Az adatátviteli protokollok, az adatok és adatbázisok minőségének megbízhatósága és szabványosítása nélkül a hálózatok, a letöltés és a szakértő rendszerek fejlesztése nem következhet be. Fontos, mégis gyakran elhanyagolt tényező az ezeknek az adatbázisoknak a kereséséhez elengedhetetlen szakértelem. Az információközvetítők azzal segíthetnek felhasználóiknak és maguknak is, hogy a rendelkezésre álló numerikus/faktografikus adatbázisokat megismerik, és felméri a várható igényeket. A

munkájuk jövőjét féltő online keresőknek ez esélyt jelent arra, hogy tudásukat erőforrásnak tekintsék, s ne mellőzzék őket.

#### Irodalom

- [1] WILLIAMS, M. E.: Computer-readable databases. 1. köt. Science, technology, medicine. Elsevier, 1985. p. VII–VIII.
- [2] Directory of Online Databases. Cuadra/Elsevier, 9. köt. 1. sz. 1988.
- [3] CODATA. Inventory of data sources in science and technology: a preliminary survey. UNESCO, 1982.
- [4] McCARTY, J.: The automated data thesaurus: a new tool for scientific information. Proceedings of the 11th International CODATA Conference. FDR, Karlsruhe, 1988.

/ALLAN, F. C.–FERREL, W. R.: Numeric databases in science and technology: An overview. = Database, 12. köt. 3. sz. 1989. p. 50–58./

(Roboz Péter)

## Teljes szövegű online adatbázisok értékelése

Az utóbbi tíz évben az adatbázisok számának rohamos növekedésével arányosan nőtt a teljes szövegű adatbázisok száma is. Használatuk azonban nem nőtt olyan gyorsan, mint a bibliográfiai adatbázisoké. Az 1970-es évek végén három pontban állapították meg a tényleges használat feltételeit:

- ▶ az információ elegendő legyen,
- ▶ időben érkezzen,
- ▶ fontos legyen a felhasználónak.

Ezek közül legalább két feltételnek kell teljesülnie ahhoz, hogy egy adatbázis sikeres legyen. Az *American Chemical Society (ACS)* szerint teljes szövegű adatbázisaira mindhárom követelmény teljesül. Az ACS 1983-ban 18 folyóiratát teljes szövegű adatbázisok formájában tette hozzáférhetővé a BRS szolgáltatóközpontnál. Ezek az adatbázisok a *Chemical Journals Online (CJO)* adatbázis részeként kerültek át 1986-ban az STN International szolgáltatóközponthoz (USA–NSZK–Japán).

Az ACS az idők folyamán egyre több folyóirat-kiadó termékeit vitte be online kereshető formában a CJO adatbázisba. Ennek eredménye, hogy a teljes szövegű információ mennyisége valóban hatalmas lett: folyóiratcikkek ezreit tartalmazza az adatbázis, némelykor nyolc évre visszamenőleg. Az információ valóban friss, ugyanis a CJO-ban szereplő folyóiratok sokszor előbb olvashatók online, mint ahogy az előfizetőkhez a lap megérkezik. Végül a teljes szövegükben online kereshető és olvasható primer folyóiratok – a felhasználók szerint – a kémiai kutatás nélkülözhetetlen segédeszközei lettek.

#### A teljes szövegű adatbázisok előállítása

A teljes szövegű adatbázis előállításának két kulcskérdése van: a nyomtatott folyóiratok átalakítása online kereshető alakra és a nagy adatmennyiség tárolása.

A nyomtatott folyóiratok előállításában használt mágnesszalagos állományokat adatbázis-formátumra kell konvertálni, ami számos részproblémát rejt magában (pl. a nem alfanumerikus karakterek kódolása). Az egyes kiadók másképp kódolják az adatelemeket, ezért azokat az adatbázis-készítés során egységesíteniük kell, ami megnöveli a költségeket.

A betöltés után a tárolás költségei igen nagyok, mert teljes szövegeket, azaz minden egyes szót és a proximity (helyzeti operátoros) kereséshez szükséges indexeket is online kereshetővé kell tenni. A CJO adatbázisban az indexállományok tárolásához háromszor akkora tárterület szükséges, mint magához a szöveghez. Ha még grafikus információt is tárolni kellene, a költségek megengedhetetlenül nagyok lennének. A tárolás költségei így is sokkal nagyobbak, mint a bibliográfiai adatbázisok esetében.

#### A felhasználók

A felhasználók döntését, hogy teljes szövegű adatbázist használjanak-e, sok tényező befolyásolja. Az egyik ilyen a *teljesség* szempontja.

A legtöbb esetben a teljes szövegű folyóiratokat 5 évnél rövidebb időre visszamenőleg tartalmazzák az adatbázisok. Ez sok keresőt arra késztet, hogy sok évre visszamenő bibliográfiai keresést végezve, a könyvtárban olvassa el a régebbi dokumentumokat.

A teljességet befolyásoló következő tényező a viszonylag kevés olyan folyóirat, amely teljes szövegben online kereshető és olvasható. A kémiában pl. mintegy 2000 folyóirat létezik, ezek közül mintegy 300 tekinthető "core" (kulcsfontosságú) lapnak. Ezeket is különböző, egymással versengő kiadók állítják elő, akik nem szívesen működnek együtt teljes szövegű online adatbázis létrehozása érdekében.

A harmadik tényező a grafikus információ, a képletek és táblázatok nehézkes online kezelése. Az adatbázisok csak minimális mértékben tartalmazznak ilyen információt, holott ezek a kutatók számára igen fontosak lennének. Igaz, vannak már újabb műszaki megoldások, amelyek az ábrák online hozzáférhetőségét is lehetővé teszik.

A felhasználói döntést befolyásolják a felhasználói szokások is. A mai felhasználók kereshetnek információt nyomtatott vagy mikrofilm-referáló folyóiratokban, CD-ROM vagy online adatbázisokban. Primer dokumentumokat kaphatnak könyvtári állományból, különnyomat-kérések útján, könyvtárközi kölcsönzéssel, újabban pedig teljes szövegű online adatbázisokból is. Az utóbbiaknak meg kell változtatniuk a korábbi felhasználói szokásokat, beidegződéseket. Sajnos, a teljes szövegű adatbázisok előállítói nem mérték fel az igényeket, és nem propagálták eléggé szolgáltatásuk előnyeit. Jelenleg még nem gazdaságos a primer folyóiratok vagy a dokumentummásolat-szolgáltatás felváltása teljes szövegű adatbázisokkal. Az előállítóknak kemény munkával kell meggyőzniük a felhasználók zömét arról, hogy mennyi előnye van a teljes szövegű adatbázisok használatának akkor is, ha nem olyan gazdaságosak, mint a hagyományos primerdokumentum-beszerzési módszerek. (Átlagban három dollár egy cikk szövege online, két dollár a másolatának beszerzése könyvtárközi kölcsönzéssel.)

A következő döntési pont: *kapcsolat a bibliográfiai adatbázissal*. A legtöbb teljes szövegű adatbázisnak van bibliográfiai megfelelője is, amely egyben a legfontosabb konkurense. A cél a kettő összekapcsolása lenne, de ma ez még alig történik meg. A *Magazine Index* már összekapcsolta bibliográfiai és teljes szövegű adatbázisát, ezáltal az előbbiben végzett keresés eredményeinek nagy része teljes szöveggé is olvasható a Dialog rendszerben. Ekkor viszont a teljes szöveg nem kereshető, szerepe csupán a dokumentum azonnali kézhezvétele.

A legtöbb esetben viszont hiányzik az együttműködés a primer és szekunder adatbázis között, pedig ezek jól kiegészíthetnék egymást (egyik a mélységet, másik a szélességet képviseli az informálódásban). Nem fejlesztenek ki megfelelő szoftvert a kétféle

állomány keresésének átfedésére. A megoldást az integrált információs rendszerek jelentenék, amelyek a rokon adatbázisokat összekapcsolják, és közös parancsnyelvvvel működnek. Egy valódi integrált rendszer a felhasználót érintő minden információt szolgáltatni tudna egy adott szakterületen: teljes cikkszöveggel, bibliográfiával, kézikönyvvel, ábrákkal, enciklopédiával. (Lásd HEARTY, J. A.-LOVE, R.: Integrated information databases: One-stop chemical research for the information specialist. = World Patent Information, 10. köt. 3. sz. 1989. p. 199-203.)

Külön kell szólnunk a *végfelhasználókról*, mint a felhasználók olyan csoportjáról, akiket a fentiekől eltérő döntési szempontok is vezérelnek. A teljes szövegű primer információ nagyrészt inkább őket vonzza, és ez a potenciális piac hatalmas lehet. Jelenleg azonban a végfelhasználók még nem mind aktív online keresők. A vállalatok és egyetemek könyvtárai nem szívesen adják át online keresésre jogosító jelszavaikat az olvasóknak. Ugyanakkor a mai aktív végfelhasználók zöme még nem kapott elegendő online oktatást, és alig tudnak valamit az online rendszerekről. Ez azonban változóban van, a mai diákok már jóval többet tudnak az online keresésről és szolgáltatásokról, mint tanáraik, sőt, egy részüknek a tantervében is szerepel az online kurzus.

A keresések egy részét közvetítők végzik, akik jól képzett, gyakorlott szakemberek. A közvetítők általában ódzkodnak a teljes szövegű adatbázisoktól. Talán félnek a hamis találatoktól, de a fő ok az, hogy nem ismerik tökéletesen a témát. Számukra is kedvezőbbé tehető a teljes szövegű adatbázisok szemináriumok, továbbképzések révén. A végfelhasználók, akik nem tapasztalt keresők, mesterséges intelligencia felhasználásával, speciális interfészekkel segíthetők át a keresés nehézségein. A végfelhasználókat segíti az adatbázisok új árpolitikája, amely inkább a "kapott értéket" fizeteti meg, mintsem az online keresési időt. Ez utóbbi ugyanis a gyakorlatlan keresők – a végfelhasználók – ellen dolgozik, míg a találatok számával arányosan növekvő ár a keresés eredményességét, nem az idejét fizeteti ki.

### Következtetések

Az adatbázis-készítők és online szolgáltatók sok pénzt fektettek abba, hogy pontosabb módszereket dolgozzanak ki a teljes szövegű adatbázisok online keresésére és megjelenítésére. Nyilván további fejlesztéseket is támogatnak, hogy az ilyen állományok hatékonyságát növeljék. A végfelhasználók is lassan, de biztosan inkább az online rendszerekhez fordulnak információs igényeik kielégítéséért. Ezek a trendek fényes jövőt jósolnak a teljes szövegű adatbázisoknak. Azt persze ne várja senki, hogy egyik napról a másikra megtérülnek a befektetések, és ugrásszerűen emelkedik az ilyen adatbázisok igénybevétele;

ehhez időre és türelemre van szükség. Az adatbázisok előállítóinak, az online szolgáltatóknak és a felhasználóknak szorosabban együtt kell működniük a siker érdekében.

/HEARTY, J. A. – ROHRBAUGH, V. K.: Current state of full text primary information online with recommendations for the future. = Online Review, 13. köt. 2. sz. 1989. p. 135–140./

(Roboz Péter)

## Nemzetközi céginformációs adatbázisok felmérése

Az üzleti információk iránti megnövekedett keresletben különleges szerepe van a céginformációs adatbázisoknak. Az e területen tevékenykedő online keresőknek már kifejlődött egy sajátos érzékük arra, hogy mely adatbázisban tudnak bizonyos gazdasági ágakban működő cégekről listát készíteni, címlistákat kinyomtatni, tulajdonviszonyokra vonatkozó információkat feltárni, vagy pénzügyi helyzetképet kapni egyes vállalatokról. Megismerték az egyes adatbázisok kedvező tulajdonságait, hiányosságait, a cégekről kapható információk mennyiségét, mélységét, a keresés árát, az adatok megbízhatóságát stb.

Régi (és jogos) panasz, hogy egészen a legutóbbi időkig, nemzetközi szinten (értsd: az USA-n kívüli országok cégeiről) nagyon nehéz online céginformációt keresni. Igaz, hogy az *International Dun's Market Identifiers* adatbázis a Dialognál 1983 óta kereshető, de csak a vezető európai cégeket tartalmazza. A kevésbé ismert európai cégekről információt legfeljebb nyomtatott formában lehet találni, ha egyáltalán létezik ilyen kiadvány. (Erre vonatkozólag – magyar szerzőktől – nagy európai helyesléssel fogadott előadás hangzott el Londonban 1988-ban [1].)

Az utóbbi években azonban bővülnek a nemzetközi céginformációs adatbázisok. Az USA-n kívüli országok adatbázisai közül főleg a *Kompass*, az *ICC* és a *Hoppenstedt* adatbázisok keltek feltűnést. A táblázatokban közölt felmérés és útmutató az ezekben és hasonló adatbázisokban való könnyebb tájékozódást kívánja segíteni. A felmérés szempontjai a következők:

- ▶ az USA-n kívüli cégek országonként,
- ▶ a cégek száma,
- ▶ az információ mennyisége és típusai.

### A felmért adatbázisok

A felmérés mindazokra az útmutató (directory) típusú céginformációs adatbázisokra kiterjed, amelyek a cégek alapadatait (név, cím, telefon, telex stb.) tartalmazzák, ezeken kívül esetleg pénzügyi információkat, létszámadatokat, mérlegadatokat stb. is szolgáltatnak. A felmérésben nem szerepelnek a speciális ágazatokra vonatkozó cégútmutatók (bank, gyógyszeripar, turizmus stb.), az előfizetést igénylő szolgáltatások, az USA cégek külföldi leányvállalatait is tartalmazó, az USA-ra vonatkozó adatbázisok, a

vállalati alapadatokat is tartalmazó, de fő hangsúllyal beruházási jelentéseket és ekonometria idősorokat közlő adatbázisok. Hiányoznak továbbá az USA-beli vállalatok mellett néhány más ország (pl. Kanada) egyes cégeit együttesen tartalmazó adatbázisok.

Az adatbázisok pillanatnyi helyzetét tükröző minden felmérés olyan, mintha egy mozgó autón szeretnénk kereket cserélni. Új adatbázisok jelennek meg, korábbiak megszűnnek vagy átalakulnak, változnak a hozzáférési feltételek.

A céginformációs útmutató típusú adatbázisok két nagy csoportja különböztethető meg:

- ▶ *Egyszerű* útmutatók, amelyek a helyi alapadatokon kívül néhány további alapadatot tartalmaznak (lásd a táblázat adatmezőit). Példa: *International Dun's Market Identifiers*.
- ▶ *Összetett* útmutatók, amelyek az alapszolgáltatásokon kívül pénzügyi adatokat, esetleg szöveges jelentést is tartalmaznak. Példa: *ICC Financial Datasheets*.

Ezekon kívül léteznek még *cégszerkezeti* útmutatók, amelyek vállalatcsaládokra és tulajdonviszonyokra vonatkozó információt tartalmaznak (példa: *Inter-Corporate Ownership*).

A felmért nemzetközi céginformációs adatbázisokat az 1. táblázat tartalmazza.

### A táblázat adatai

**Tartalom (és forrás).** Az adatbázisba felvett vállalatok száma, amely egyben az adatbázis méretének (és típusának) mércéje is. Fontos tudni, hogy a rekordok száma nemcsak azzal változik, hogy az adatbázis bővül, hanem szolgáltatóként is. Zárójelben van feltüntetve az adatbázisban szereplő információ forrása.

**Online szolgáltatóközpont(ok).** Azok a szolgáltatók, amelyek az USA-ból nem érhetők el közvetlenül, nincsenek a táblázatban feltüntetve. (Azok sem, amelyek Magyarországról nem érhetők el. – A ref.)

**Adatmezők.** A rekordok adatmezőinek nem teljes listája áll ebben a rovatban. Egyes adatbázisoknak számos kereshető (esetleg több mint száz kiíratható) adatmezője is lehet. Ebben az oszlopban az egyedül és a különösen hasznos mezők vannak feltüntetve, de pl. a cím- és cégnévadatok, telefonszámok stb. nem, mert ezek minden adatbázisban meglévő adatmezők. A kereshető adatmezők szolgáltatóként is változnak.