

leíró kifejezések nem eredményezik a keresett válasz elemeit. Ekkor minden fogalom szemantikai környezetét meg kell határozni. Igaz, hogy egy adott témakör szakértője képes fejben azonnal összeállítani egy "rögtönzött teauruszt", amelynek azonban az adatbázis által használt szógyűjtemények is meg kell felelnie. A nem szakértő ugyanakkor különböző segédeszközöket vehet igénybe a probléma megoldására.

Az adatbázisok előállítói által készített *teauruszok* pótolhatják valamelyest a felhasználó hiányos ismereteit. Különösen a kémiai, vegyipari adatbázisok teauruszai hatékonyak e tekintetben, s adott vegyület struktúrájának kanonikus megjelenítése megoldotta a nomenklatura gondját a vegyészek számára számos adatbázis esetében.

A közvetett eszközök közé tartoznak a *konkordancia-listák* (a legfontosabb kulcsszavak más kulcsszavakkal együttes előfordulásait rögzítő fájlok) és a *hivatkozási fájlok* (új kifejezések, új szóhasználat feldejtéséhez hasznosak).

A fogalmi megközelítéssel az adott területre szakosodott végfelhasználók nagyobb eredmény reményével foghatnak a kereséshez, de az adatbázisok használatának szakembere a segédeszközök ötletgazdag alkalmazásával nagyobb hatékonyságot tud elérni.

A "Máté-effektus" avagy a rendszerszakértő

A jó minőség/ár hányadost képviselő keresés paraméterei többdimenziósak és változóak. Csak az adatbankok és adatbázisok keresésének szakértői ismerik ezeket kellő mértékben, ill. csak ezek a szakértők tudják kellőképpen hasznosítani e paramétereket. Az alkalmi felhasználó – bármennyit tud is az illető szakterületről – nem képes hasonló teljesítményre. Csak a legalább 100 adatbázisban

keresést végzett szakember képes a legígéretesebb fájlok kiválasztására, ill. az eredmények kiegészítésére a kevésbé kézenfekvő fájlokból. A keresési szakértő gyorsabban reagál, kifinomultabb stratégiát alkalmaz és még a legegyszerűbb esetben is 25%-kal gyorsabban és sokkal jobb eredményt ér el a szoftver tökéletesebb ismerete révén.

A "Máté-effektus" ("Mert annak, akinek van, még adnak, hogy bőven legyen neki; akinek meg nincs, attól még amije van is, elveszik" – olvasható Máté evangéliumában) ebben a vonatkozásban azt jelenti, hogy a magas színvonalú információkereséseket végző szakemberhez mind gyakrabban fordulnak. Következésképpen egyre tapasztaltabbá és eredményesebbé válik, ezért több keresést bízunk rá, s egyre jobban végzi ezt a tevékenységet. (A "Máté-effektust" a szakirodalom eddig az "információban szegények" és az "információban gazdagok" közötti különbség leírására használta. – A ref.)

Az eseti felhasználó – aki eredményeit a ráfordított idő, a keresés eredményeként nyert válaszok és a kapott számlák alapján értékeli – rájön arra, hogy jobban jár, ha a keresést a szakemberre bizza. Mindenesetre ezt látszanak bizonyítani az *Institut Français du Pétrole* online kereséssel kapcsolatos tapasztalatai. A mennyiségét és komplexitását tekintve növekvő kereslet, az egyre komplexebbé váló fájlok és a kötött költségvetés körülményei között csak így lehetett a szolgáltatást az igényeknek megfelelően üzemeltetni az intézménynél. Ehhez azonban centralizált szervezetre van szükség a teljesítmények és a költségek figyelemmel kísérése érdekében.

/MOUREAU, M.: Cost and know-how – the "Matthew effect" in information retrieval. = Online Review, 11. köt. 6. sz. 1987. p. 355–360./

(Hegedűs Péter)

Bevezetés a numerikus adatbázisok világába

Sokan közülünk félnek a számoktól. Az információ szakembernek azonban nem szabad ebbe a csapdába esnie. A numerikus adatokkal is kényelmesen és eredményesen dolgozhatunk, anélkül, hogy a számok szakemberei lennénk. A numerikus adatbázisokat arra tervezték, hogy nagy adatmennyiségeket lehessen velük hatékonyan kezelni, használatukat pedig semmivel sem nehezebb megtanulni, mint a bibliográfiai adatbázisokét.

Mik azok a numerikus adatbázisok?

A numerikus adatbázisok a nem bibliográfiai (vagy forrás-) adatbázisok körébe tartoznak. Az ilyen adatbázisok magát a keresett információt tartalmazzák, míg a bibliográfiai adatbázisok azt, hogy az információ

hol lelhető fel*. A nem bibliográfiai típusú adatbázisok körébe tartoznak a teljes szövegű adatbázisok (pl. az **INVESTEXT**), az útmutatók (pl. az **Electronic Yellow Pages**), a grafikus adatbázisok (pl. a **Chemical Abstracts** szerkezetiképlet-állománya), valamint tárgyunk, a numerikus adatbázisok.

A tisztán numerikus adatbázisok, amelyekkel foglalkozni akarunk, a számadatok mellett csak nagyon kevés – meghatározás jellegű – szöveges információt tartalmaznak. Ezek az adatbázisok elsősorban a

* Az elhatárolás nem mindig ilyen egyszerű. Egyes adatbázisok, mint például a **Predicasts PROMT** és a **Metals Data File** együtt tartalmazzák magukat a numerikus adatokat és forrásuk bibliográfiai adatait. – A ref.

természet- és társadalomtudományok körében használatosak. Mi most legelterjedtebb változatokra, a gazdasági-üzleti adatbázisokra koncentrálunk.

A numerikus gazdasági-üzleti adatbázisok száma az utóbbi években rohamosan nő. Forrásukul szolgálhatnak állami szervek, nemzetközi szervezetek vagy magáncégek adatai, sőt egyéni kutatók adatai is. Az adatok eredetét és esetleges feldolgozásuk módját az adatbázis dokumentációja írja le, és ezt jól meg kell értenünk, ha helyesen akarjuk használni az adatbázist.

Hol található a numerikus adatbázisok?

Egyes numerikus adatbázisok a nyilvános szolgáltatóközpontoknál érhetők el (ilyen központok például a *DIALOG* és az *I. P. SHARP*), mások magánál az előállítónál (ilyen adatbázis például a *U. S. Department of Commerce Economic Bulletin Board*), megint mások a felhasználó intézmény saját rendszerén történő futtatásra szerezhetők be. Az utóbbi típusú adatbázisok használata helyről helyre nagyon erősen változik.

Az előállítónál elérhető adatbázisok többségükben nagyon egyszerű menürendszerben használhatók, amit sok online segédinformáció könnyít meg.

A nyilvános szolgáltatóközpontok két csoportba oszlanak, a szöveges és a numerikus rendszerekre.

A szöveges rendszerekben (pl. a *DIALOG*) bibliográfiai, teljes szövegű és útmutató típusú adatbázisok mellett találhatóak a numerikus adatbázisok. Az adatok keresésére ugyanaz a rendszer szolgál, mint a szöveges adatbázisok keresésére. Ha ezt a keresőrendszert megtanultuk, már megvannak az alapismerteink a numerikus adatbázisok használatához is.

A numerikus szolgáltatórendszerek (pl. az *I. P. SHARP*) parancsnyelve más stílusú, általában sokkal bonyolultabb, és sokkal több bemenő információt igényelnek a felhasználótól. Használatukhoz sokkal több tanulásra és gyakorlásra van szükségünk, mint a szöveges rendszerek használatához. Ez azonban egyúttal nagyobb rugalmasságot nyújt az adatvisszakeresésben és az adatok alakításában, valamint nagyobb uralmat az információ felett.

A következőkben a nyilvános szolgáltatórendszereken elérhető adatbázisokra összpontosítunk. Ne felejtsük el azonban, hogy sok hasonló vonás van a különböző helyeken elérhető numerikus adatbázisok használatában.

Az adatok

Hogy a numerikus adatbázisokat használjuk, ahhoz éppúgy nem kell statisztikusoknak lennünk, ahogy nem kell pszichológusoknak lennünk a *Psyc-Info* adatbázis használatához. A legfontosabb, hogy megértsük azokat a változókat (deskriptorokat), amelyekre ezek az adatbázisok használnak. Általában négy alapváltozó használatos.

A **típus** változó írja le, hogy mire vonatkoznak az adatok. Egy kamatláb-adatbázis például tartalmazhatja az elsődleges kamatlábakat, a jelzáloghitelek

kamatlábait, a kereskedelmi hitelek kamatlábait stb. Egyes adatbázisok csak egyféle típusú adatokat tartalmaznak, mások sokfélélt.

Az **időbázis** változó írja le egyrészt azt, hogy mekkora időintervallumra vonatkozik egy-egy adat (pl. napi, heti, havi, negyedéves, éves adatok), másrészt azt, hogy mekkora időszakot fed le az adatbázis (pl. az *I. P. SHARP* rendszerben a **Consumer Price Index** adatbázis 1914-től napjainkig, ugyanez a *DIALOG*-ban a legutóbbi 20 évet). Egyazon típusú adat többféle időintervallumra is szerepelhet az adatbázisban, pl. havi és éves adatként. Egyes szolgáltatásrendszerekben manipulálhatók is az adatok, például a havi adatok összevonhatók negyedéves adatokká.

A **mértékegység** változó adja meg, hogyan kell értenünk az adatot (pl. dollárban, százalékban, kilogrammban). Gyakran egyazon mennyiség több mértékegységben is megtalálható az adatbázisban.

A **részletezés szintje** adja meg például, hogy milyen földrajzi egységre, korosztálycsoportra, iparágra vonatkoznak az adatok. Ez a változó gyakran meghatározó abból a szempontból, hogy melyik adatbázist válasszuk.

A négy alapváltozó mellett még néhány további fogalommal is tisztában kell lennünk a numerikus adatbázisok használatához.

Az **indexszámok** valamely meghatározott bázisidőszakhoz képest a változást írják le. A **Consumer Price Index** adatbázisban például 1967 bázisév, ezt veszik 100-nak. Ha az index 105, az 5%-os növekedést jelent 1967-hez képest.

Az **idősor** valamilyen mennyiség egymás utáni időszakokra vonatkozó adatait tartalmazza (pl. a fogyasztói árindex változása évenként).

A **keresztmetszet** egyazon időszakra vonatkozó különböző adatok halmazát jelenti. Célja általában különböző adattípusok összehasonlítása egy bizonyos időpontban (pl. a munkabérek és ledolgozott órák 1985 augusztusában).

A **szezonális korrekció** arra szolgál, hogy kiiktassuk az adatokból a szezonális jelenségek (pl. az időjárás, a munkanapok eltérő száma) hatását.

A **regresszióanalízis** olyan matematikai statisztikai eljárás, amellyel kapcsolatot határozhatunk meg két vagy több változó között. Előrejelzésre is használható.

További változókat az adatbázisok dokumentációjából ismerhetünk meg. Nagyon hasznos, ha kezünk ügyében van a közgazdasági és statisztikai fogalmak valamilyen kézikönyve is. Ilyen lehet például a *McGraw-Hill Dictionary of Modern Economics* (3. kiadása: McGraw-Hill, New York, 1983).

A numerikus adatok lehívása

Szándékosan használjuk a lehívás vagy visszakeresés szót, nem pedig a keresést. A numerikus adatbázisokban ugyanis – ritka kivételektől eltekintve – magukra a számokra nem lehet keresni. Nem kérdezhetjük például, hogy mikor és hol volt 10% a kamatláb. Általában a változókat kell megadnunk, hogy megkapjuk a nekik megfelelő adatokat. Kérdez-

hetjük például a jelzőlogóhitelek és a takarékbetétek (típus) havi kamatlábát az elmúlt évre (időbázis) az Egyesült Államokban (a részletezés szintje). Ehhez előre kell tudnunk (a dokumentációból), hogy ezek a változók szerepelnek-e az adatbázisban. Vagyis előre tudunk mindent az adatról, kivéve magát az értéket. Ez az érték az, amit lehvívunk vagy visszakeresünk az adatbázisból.

Egyes keresőrendszerek (DIALOG) lehetővé teszik a változókra való szöveges keresést is. Ilyenkor azonban sok nehézséget és pontatlanságot okoz a lehetséges többféle írásmód, a különböző rövidítések, a többféle lehetséges megnevezés stb. Célszerű ezért inkább az adatbázis *kódrendszerével* végezni az adatlelvívást. Ilyen kódrésze a legtöbb numerikus adatbázisnak van a különböző változókra. Használatuk növeli a pontosságot és időt takarít meg.

A végső lépés a visszakeresett adatok alakítása és megjelenítése. Egyes rendszereknek (I. P. SHARP) az adatok online manipulálására is vannak eszközeik. Más rendszerekben csak arra van lehetőségünk, hogy a saját mikroszámítógépünkbe letöltve az adatokat, azokat ott utólag, offline módon alakítsuk az igényeinknek megfelelő formára.

A numerikus adatbázisok használata során alapvetően fontos a dokumentáció igénybevétele, mert az adatok nem magyarázzák önmagukat. Ne várjunk azonban segítséget a dokumentációtól a visszakeresett adatok *interpretálásában*. Azt például nem találjuk meg ott, hogy mi az adatokban látható infláció oka.

A szöveges szolgáltatórendszerekben megszoktuk, hogy a költségek a kapcsolati időtől és a kinyomtatott eredmény terjedelmétől függnék. A numerikus szolgáltatórendszerekben azonban gyakran ettől eltérő árképzéssel találkozunk: a költség a számítógép központi egységének a használati idejétől függ. Ilyen rendszerben tehát lehet, hogy egy

terjedelmes eredményt adó egyszerű adatlelvívás viszonylag olcsó, míg egy kevés eredményt adó, de bonyolult feldolgozással járó munka meglepően drága. Amíg kevés a tapasztalatunk, nagyon nehezen becsüljük meg a várható költséget.

Ha más számára végezzük az információkeresést, fontos az alapos előzetes interjú. Ha az ügyfél számkra idegen terminológiát használ, világosan tisztáznunk kell, hogy mire gondol.

Miért van szükségünk numerikus adatbázisokra?

A numerikus adatbázisok lényegében nem tartalmaznak több adatot, mint a nyomtatott kézikönyvek, használatuk megtanulása pedig kétségtelenül többlet-erőfeszítést igényel. Miért van mégis szükségünk rájuk?

A számítógépek az adatbázisok gyakori és gyors aktualizálását teszik lehetővé. Az adatbázisok ezért mindig sokkal frissebb adatokat tartalmazhatnak, mint nyomtatott megfelelőik. Kiküszöbölhető velük az adatok lemásolása során fenyegető elírások veszélye. Nagy időmegtakarítást nyújt a számítógép az adatok manipulálásában, átalakításában. Végül, a számítógépek óriási tárolókapacitása hatalmas adattöbbleteket gyors elérésére ad módot.

Az adatok mennyisége nemcsak óriási, de nagyon gyorsan növekszik is. Már nemcsak az adatok hagyományos keresése válik egyre lehetetlenebbé, hanem az is, hogy az egyre több szükséges adatot egyáltalán kinyomtassuk. A kiutat a számítógépes adatbázisok jelentik.

/SUOZZI, P.: *By the numbers: an introduction to numeric databases.* = *Database*, 11. köt. 1. sz. 1987. p. 15–22./

(Válas György)

Információs hálózatok nyelvi eszközei és kompatibilitási problémái

Az információs hálózatok típusai és a nyelvi kompatibilitás

A hálózatok *deduktív* létrehozásakor már az adatbázisok tervezésénél figyelembe veszik a nyelvi kompatibilitás követelményét: a hálózatban összekötött adatbázisok egységes, vagy legalább összehangolt elvek szerint felépülő nyelvi eszközöket használnak. Gyakoribb azonban, hogy az információs hálózatok *induktív* módon jönnek létre, vagyis már működő információs rendszereket kapcsolnak össze. E rendszerek feldolgozási módszere és technológiája eltérő. Az ilyen típusú hálózatokban a nyelvi kompatibilitást radikális átalakítás nélkül kell elérni olyan mechanizmus segítségével, amely elvégzi a fordítást az adatbázist kibocsátó nyelvről a hálózat szolgáltatási

nyelvére. Hasonló problémák merülnek fel a hálózat szempontjából külső (pl. külföldi) adatbázisok átvételekor.

A külföldi adatbázisok információkereső nyelveinek struktúrája igen különböző. Mivel a nyelvi eszközök egységesítése ez esetben lehetetlen, a hálózatoknak olyan univerzális mechanizmusra van szüksége, amely képes átalakítani bármely információkereső nyelvet a befogadó rendszer (hálózat) nyelvére.

A külföldi hálózatok és kereskedelmi információs szolgáltatások (pl. Dialog, Data-Star, SCD stb.) adottan veszik az adatbázisok információkereső nyelvére.