

- El kell törölni a szabad piac versenyességét tönkretévő, indokolhatatlan különadókat, amelyek a nem hazai információforrásokat és információkezelő eszközöket terhelik.

Utópia-e a szabad információáramlás?

A vázolt célok megvalósítása rengeteg időt, pénzt és szellemi befektetést igényel, mindezt nemzetközi szinten. Ugyanakkor magának a feladatnak felbecsülhetetlen értéke van: amíg ezeket a követelményeket legyőzendő akadályoknak tekintjük, nem nyomhat agyon bennünket a tudományos és műszaki információk gigantikus tömege. A jelzett területeken gyakran

dominálnak a politikai-gazdasági megfontolások. Egyenrangú partneri viszonyban kell együttműködnünk a problémák megoldása érdekében, hogy a globális szabad információáramlás ne maradjon megvalósíthatatlan utópia.

/LUCKENBACH, R.: The free flow of information: a utopia? Ways to improve scientific and technological information and its international exchange. = Journal of Chemical Information and Computer Sciences, 28. köt. 2. sz. 1988. p. 94–99./

(Szöllősy Éva)

Az adatbázisok pontossága. Egy felmérés eredményei

A nyugati gazdasági szakemberek és döntéshozók nemcsak belső, hanem külső adatbázisokat is használnak munkájuk, különösen prognóziskészítő tevékenységük közben. Míg az előbbieket paramétereit ismerik, az utóbbiakét nem. Pedig munkájuk sikere erősen függ ezeknek a külső gazdasági (business) adatbázisoknak a minőségétől, az általuk szolgáltatott adatok pontosságától.

Ezért volt indokolt az az *adatbázis-szolgáltatók* (vendors) körében végzett felmérés, amely a következőkre terjedt ki:

- az adatgyűjtési módszerek,
- a gyűjtött adatok típusai,
- az adatokhoz való hozzáférés formátumai,
- az adatok pontosságának ellenőrzési, karbantartási módszerei és mások.

A felmérés módszere

A felmérés készítői mindenekelőtt annak felderítésére koncentráltak, hogy az adatbázis-szolgáltatók miképpen biztosítják – ellenőrzik, ill. növelik – az *adatok pontosságát*. Ezért számos változót vizsgáltak. Például: az adatbázis nyers vagy számított adatokat tartalmaz-e; az adatbevitel miképpen és milyen gyakorisággal megy végbe; milyen ellenőrzési módszerekkel és gyakorisággal élnek a pontosság garantálása érdekében; melyek a hibák elsődleges forrásai; a hibák felfedése után milyen a kiküszöbölés módszere és átfutása.

Az adatbázis-szolgáltatók katalógusaiból összesen 185 olyan szolgáltató szervezetet azonosítottak, amelynek adatai alkalmasak a gazdasági előrejelzésre. Ennek a 137 USA-beli és 48 USA-n kívüli szervezetnek küldték ki az imént felsorolt változók iránt érdeklődő kérdőíveket. A kérdőíveket – nemegyszer sürgetésre – a megkeresettek 31%-a juttatta vissza többé-kevésbé érdemi válaszokkal.

Az eredmények

A választ adó szolgáltatók 94%-a forrásadatbázisokat kínál, vagyis olyanokat, amelyek numerikus, szöveges-numerikus tényadatokat vagy/és teljes szövegű információkat tartalmaznak. Az adatbázisok zöme tehát *mennyiségi* prognózisok készítésére használható fel. A szolgáltatók 28%-ának van bibliográfiai adatbázisa (is). Ezeket minőségi és műszaki előrejelzésekre lehet igénybe venni.

A válaszoló szervezetek nagy része kettőnél több adatbázist szolgáltat. Az adatbázisokban a változók (paraméterek) száma 8-tól néhány ezerig terjed; a legtöbb adatbázisban a változók százaléka kereshető online módon. A gazdasági adatok között árindek, részvény-, demográfiai, piaci részesedési, export-, importadatok, továbbá a nemzeti össztermékre, a munkaerő összetételére, a szabadalmakra, védjegyekre stb. vonatkozó adatok találhatóak.

A felhasználónak mindenesetre gondosan meg kell vizsgálnia, hogy mely adatbázisok felelnek meg leginkább az igényeinek, mert a változók és a specifikus adatok száma akkor is módosulhat, ha az adatbázis történetesen csak egy bizonyos szakterületet fed le. Fontos felhasználói szempont az is, hogy milyen formában kapja meg az adatokat a szolgáltatótól. A prognóziskészítő szempontjából különös jelentősége van annak, hogy mennyire összesített numerikus adatokat várhat el egy-egy adatbázistól, vagy éppen ellenkezőleg, összesítetlen, nyers adatokat kereshet-e benne (ezek regionális vagy helyi szinten többet érnek, mint az összesített adatok).

A válaszolók csaknem háromnegyed része nyers numerikus adatokat kínál. A felmérésbe bevont szolgáltatók 30%-a grafikus információkat is nyújt. Jól szervezett statisztikákat a szervezetek 58%-a, kész előrejelzéseket pedig 10%-a közöl.

Az adatok kiválasztási kritériumai között a legfontosabb a forrás hozzáférhetősége és megbízhatósága. Az adatok teljességében az állami (szövetségi) és a vállalati forrásokra való támaszkodás esetén nagyrészt meg lehet bízni, ám a nemzetközi adatok teljességében már korántsem.

Az adatbázis-szolgáltatók 28%-a azonnal beviszi a friss adatokat, 47%-a valamilyen periodikusságú kötegelt adatbevittet alkalmaz. A bevitt kötegelt módja hátrányosan befolyásolja a naprakészséget, ezért az azonnal "cselekvők" 28%-a meglehetősen kevés. A felhasználónak, különösen ha előrejelzést készít, a kötegelt feldolgozású adatbázisok használatakor ismernie kell az utolsó aktualizálás időpontját, nehogy "megjárja".

Az *adathibák* túlnyomó része átírási/beviteli hiba. A második legnagyobb hibacsoport a forrás megbízhatatlanságából adódott.

A felmérésben szereplő szolgáltatók csaknem 20%-a nem ellenőrzi adatainak pontosságát (1. táblázat). A legtöbben nem foglalkoztatnak adatelemzőket vagy lektorokat, nem használnak matematikai vagy statisztikus ellenőrzési módszereket sem.

1. táblázat

A pontosság ellenőrzésére alkalmazott módszerek*

Módszer	A válaszolók száma = 43	
	Szám	Százalék
Nem ellenőrzi a pontosságot	8	18,6
Véletlen mintavételes ellenőrzés	14	32,5
Primer adatforrás visszakérdezése	18	41,9
Teljes adatellenőrzés	13	30,2
Lektor/elemző szakember	2	4,7
Matematikai összefüggéseket, szórásszámítást alkalmazó ellenőrzési módszerek	1	2,3

* Számos szervezet több módszert is alkalmaz.

Ha valamely hibára fény derül, a szolgáltatók 48%-a alkalmaz átfogó ellenőrzést. 24%-uk azonban ilyenkor sem néz utána rendszeresen az adatok pontosságának. A válaszadók 44%-a nem jelentette, milyen gyakran ellenőrzi az adatokat, 14%-a havi vagy ennél ritkább ellenőrzést alkalmaz. A szolgáltatók fele, ha kell, azonnal javítja az adatokat, ami legtöbbször 48 órán belüli javítást jelent. Tehát nem elhanyagolható veszély, hogy valamely előrejelzés hibás adatokra épül, különösen akkor, ha valahol kéthetenként vagy még gyakrabban készítenek modellt vagy prognózist.

A szélsőséges adatok kezelésének módjáról készült felmérés eredményeit mutatja a 2. táblázat. A szélsőséges adatok ugyanis igen nagy problémát

jelentenek az előrejelzésben, mert többnyire érthetetlenek, ill. nehezen magyarázhatók. Az előrejelzés szempontjából az ilyen adatokat úgy kell közölni, amilyenek, feltéve, hogy helyességüket ellenőrizték. A többi már a prognóziskészítőre tartozik.

2. táblázat

A szélsőséges adatok kezelésének módszerei

Módszer	A válaszolók száma = 47	
	Szám	Százalék
Feldolgozás "olyan, amilyen" alapon	18	38,3
Törlés az adatok közül	10	21,3
Statisztikai kiigazítás	15	31,9
Adatmegerősítés a forrásnál az azonnali korrekció érdekében	1	2,1
Különféle összetett módszerek	3	6,4

Következtetések

A felmérés eredménye számos olyan kérdést vet fel, amely mind a szolgáltatók, mind a felhasználók számára fontos. Az előbbieket számára az adatforrások kiválasztási kritériumaira, a pontosság ellenőrzésére és a szélsőséges adatok kezelésére vonatkozóan kínál megszívlelendő tanulságokat. A szolgáltatók a források megbízhatóságának kiemelkedő fontosságát egyébként meg is értik. A felhasználók szemszögéből sok minden tehető az adatbázisok minőségének javítása érdekében, ha fennáll a kommunikáció és az együttműködés a szolgáltató és ügyfelei között. Például elektronikus üzenetközvetítéssel a felhasználó tudomására lehet hozni az aktualizálás idejét és az adatok kezelésének sajátosságait (szélsőséges adatok kezelését stb.).

A felhasználók – e felmérés tanulságait hasznosítva – még gondosabban járjanak el a szolgáltatók által feldolgozott adatforrások értékelésében, az alá- vagy túlbecslés megállapításában. Ez lehetőséget ad a modellkészítőknek arra, hogy megfelelő korrekciókat alkalmazzanak.

Az adatbázisok pontosságának további vizsgálatát e felmérés módszerén kívül objektív módszerekkel is tanácsos elvégezni, különösen az olyan jellemző változók esetén, mint amilyenek az árindexek, a demográfiai adatok stb. Az adatbázisok pontossága meghatározó követelmény, mert minden sikeres tervezés, megalapozott gazdasági döntés a helytálló prognózisokon múlik, s ezek pedig a pontos adatszolgáltatástól függenek.

/MAHMOUD, E. – RICE, G.: Database accuracy: results from a survey of database vendors. = Information Management, 15. köt. 5. sz. 1988. p. 243–250./

(Roboz Péter)