

való áttéréssel, a kutatások pályázati kiválasztásának a bevezetésével az alapvetési programok tervezésénél megnő a tudományos tematika iránti igény. A kutatás minőségének objektív szakértői értékelése az aktuális információs helyzet mélyreható tanulmányozása nélkül lehetetlen. Ehhez a korszerű tudományos-műszaki haladás színvonaláról, a tudományban fellelhető részokről és szűk keresztmetszetekről, a kutatás tárgyának perspektíváiról és fejlődési tendenciáiról stb. szóló információk kellenek. Ismerni kell a hasonló kutatásokat folytató hazai és külföldi szakemberek által elért eredményeket, továbbá azt, mely kutatóközpontok és személy szerint ki folytat hasonló kutatást, ki játszik vezető szerepet a probléma kidolgozásában.

Az ilyen információk előállítására legyen a fő helyen az akadémiai információs központ tevékenységében. Ugyanakkor arra kell törekedni, hogy a létrehozott dokumentumok ne csak az anyaintézménynek legyenek érdekesek, hanem keressék őket a világpiacra is.

A második irány teljesen új. Alapját az az 1988. évi rendelet képezi, amely szerint az információs részlegnek közvetítői feladatokat kell ellátniuk a tudományos eredmények népgazdasági hasznosítását illetően. E rendelet megvalósításához meg kell szervezni a befejezett KF-munkák vagy szakaszai eredményeiről szóló információk gyűjtését, feldolgozását, rendszerezését és értékelését; rendszeresen kutatni kell az innovációk iránti igényeket; aktívan reklámozni kell az UTA intézményeinek tudományos eredményeit; el kell látni az újdonságok bevezetésére vonatkozó szerződésekkel kapcsolatos teendőket.

Az UTA-ban már megtörténtek az első lépések a fenti irányok fejlesztésében. Számos intézményben működik olyan automatizált információs rendszer, amely hatékony információkeresést tesz lehetővé a kutatások tervezése és végzése során a megalapozott döntések előkészítéséhez. Megkezdődtek a közvetítői feladatokkal kapcsolatos munkák. Az UTA tudományos információs osztályában kidolgozták a bevezetésre alkalmas újdonságokról szóló adatok gyűjtésének a módszertanát. Adatbank épül azoknak a KF-munkáknak az eredményeiről, amelyeket kiállításokon, vásárokon lehet bemutatni. Automatizált rendszerben gyűjtik azokat az adatokat, amelyek alkalmasak az UTA intézményeiben elért tudományos eredményeknek és a termelés innovációs igényeinek az összehasonlítására.

Kidolgozás alatt áll a kereskedelmi alapokon működő *Ukrakademinform* integrált rendszerének a koncepciója. E rendszer egyesíteni fogja az egyelőre még különálló elemeket: a tudományos intézmények problémára orientált adatbankjait, az innovációs alapot, a reklám- és szervizszolgáltatást stb. Az egységes infrastruktúra célja az UTA információs erőforrásainak és az egyes intézmények tudományos eredményeinek a lehető legteljesebb kihasználása.

/BELAĀ, A. A. – KOLTUN, A. Ā. – ŠIROKOV, V. A.: Naucno-informacionnaā deatel'nost' v AN USSR: problema perestrojki. = Naucno-tehničeskaā informaciā, Ser.1. 1990. 5. sz. p. 2–5./

(Viszocsekné Péteri Éva)

## Az elektronikus információ és a nyelvi nehézségek

A nyilvános adatbázisokban tárolt információkhoz a távolságoktól és országhatároktól függetlenül bárki hozzáférhet. Az online rendszerekhez nemzetközi és nemzeti hálózatok útján kapcsolódhatunk, a CD-ROM adatbázisok hordozható közegen tárolnak nagy mennyiségű információt.

Az információhoz való szabad hozzáférés ellenére az online és a CD-ROM információkeresést nehezítik: a kereső- (parancs-) nyelvek sokasága, az alkalmi vagy a kevés tapasztalattal bíró keresőket sújtó online árendszer, a hardver hiánya, az infrastruktúra (pl. távközlési kapcsolatok) elégtelensége, a pénzhiány stb. Sok erőfeszítés történik eme akadályok megszüntetésére, de furcsa módon, az elektronikus információ nemzetközivé tétele folyamán az egyik probléma megoldása egy másik akadályt erősít. Ilyen nehézség a kommunikáció nyelve, a természetes nyelv, amellyel az információt adó fél az információt fogyasztó féllel kommunikál. A kommunikáció csak akkor válik lehetővé, ha az adó és a vevő egy nyelvet

használ. Emiatt a nyelv inkább gátat jelent, mint kommunikációs csatornát a nemzetközi elektronikus információ terén.

### A nyelvi akadály

A nyelvi nehézség a publikált tudományos és műszaki információval kezdődik. Az angol nyelvű publikált szakirodalom részesedése diszciplinától függően 50% és 90% között van, a többi főleg orosz, német, francia és japán nyelvű közlemény. A számítógéppel olvasható információk terén a nyelvi akadály ugyanúgy előtérbe kerül, mint a nyomtatott szakirodalom esetében; ha nem jobban. Az online és a CD-ROM adatbázisok mintegy 20 nyelven készülnek. Senkitől, még a közvetítőktől sem kívánhatjuk meg, hogy ezeknek akár a töredékén is tudjanak információt keresni és értelmezni.

A nyelvi akadály az online kereshető információ hasznosításában a felhasználó földrajzi helyétől vagy anyanyelvétől függetlenül létezik. Ez nem azt jelenti,

hogyan az akadály egyenlő mértékben áll fenn mind-egyik nyelvészeti csoportban. Az angol anyanyelvű, vagy az angol nyelvet jól tudó kereső személyek több adatbázishoz vagy szolgáltatóhoz képesek – nyelvi szempontból – hozzáférni, mint a nem angol munkanyelvű keresők. Másrészt viszont az angol nyelvű keresők, akik nem híresek nyelvi sokoldalúságukról, szembetalálják magukat a nyelvi akadállyal, amint egy nem angol adatbázist kellene használniuk.

Az Európai Közösség Bizottsága (*Committee of the European Communities = CEC*) igen sokat tesz a nyelvi sorompók eltávolítása érdekében. 1988-ban az 500 információfelhasználó körében folytatott felmérés szerint legtöbbszörnek ugyan nem volt nyelvi nehézségük, ha saját országbeli online szolgáltatásról volt szó, de egyharmaduknak már gondot okozott, ha egy Nyugat-Európában másutt hozzáférhető szolgáltatást akart használni. Ez a szám egyharmadról a felhasználók felére nő, ha Nyugat-Európán kívüli szolgáltatókkal kerül szembe. Három fő akadályt említenek: a releváns információk forrásának felderítését, a nem tökéletes dokumentációt és a nyelvi nehézséget. 42%-ban a szolgáltatóközpont, 53%-ban az adatbázis használatakor jelent fő akadályt az idegen nyelv, holott a megkérdezettek 70%-a végez online keresést számára idegen nyelven.

Az észak-amerikai információkeresők kevésbé érzékelik a nyelvi nehézségeket, mint az európaiak, sok angol nyelvű szolgáltatóközpontjuknak (host) és adatbázisuknak köszönhetően. Az USA-ban készülő online adatbázisok részaránya 56%, az EGK országai-ban 27%, a világ többi részén 17%. Az EGK adatbázisainak nagy része is az Egyesült Királyságból, ugyancsak angol nyelvű országból származik. Az arány még inkább eltolódik a CD-ROM adatbázisoknál: 1988-ban 78%-uk származott az USA-ból, 14%-uk az EGK országaiból, és csak 8%-uk máshonnan.

#### Az adatbázisok eloszlása nyelvek szerint

A *Computer-Readable Databases* címen megjelent, a Dialog hostnál online is hozzáférhető adatbázis-katalógus szerint az adatbázisok túlnyomó többsége, 85%-a angol nyelvű. Egy másik adatbázis-útmutató, a Data-Staron kereshető (de nyomtatásban is kapható) *Cuadra Directory* 1988-as adatai is láthatók az 1. táblázatban; ez utóbbiban azonban csak az online hozzáférhető adatbázisok vannak meg, a floppylemezen vagy CD-ROM hordozón kapható adatbázisok nélkül. (A Cuadra 1990-es kiadásában már ezek is benne vannak. – A ref.)

A táblázat szerint az angolon kívül csak a francia, a német, a spanyol, a japán és a svéd nyelvű adatbázisok részesedése haladja meg az 1%-ot; ezek és az angol együttesen alkotják az adatbázisok 97%-át. A nyelvek rangsorolása a kétféle kiadvány szerint csaknem azonos, csak az ötödik és hatodik nyelv cserél helyet. A Cuadrában a japán kivételével valamennyi nyelvből több adatbázis van, mint a Computer-

Readable útmutatóban (pl. spanyolból csaknem háromszor annyi), de mindkettőben magasan vezet az angol.

#### 1. táblázat

##### Az adatbázisok nyelvek szerinti megoszlása

Nyelv	Adatbázisok	
	Computer-Readable útmutató adatai	Cuadra
Angol	3171	2616
Francia	205	335
Német	139	182
Spanyol	47	131
Japán	39	37
Svéd	29	63
Egyéb	110	*

\* Ismeretlen

A 2. táblázatban az adatbázisok típusainak nyelvi eloszlása látható, a Computer-Readable Databases besorolása szerint: bibliográfiai, útmutató, teljes szövegű, numerikus, statisztikai, grafikus, hirdetőtábla és tranzakciós adatbázisok. Az adatbázisok egy része egynél több típusba is besorolható, mint ahogy a többnyelvű adatbázisok is mindegyik nyelvnél fel vannak tüntetve. Zárójelben az előfordulás százaléka olvasható. A numerikus és teljes szövegű faktografikus adatbázisok között az angol nyelvűek részaránya nagyobb, mint a bibliográfiai adatbázisok között.

A hordozható adatbázisok új Cuadra-útmutatójában, az 1989-es kiadásban 409 adatbázis szerepel, köztük 333 CD-ROM adatbázis (számuk 1990-re meghaladta a 800-at. – A ref.). Ezek nyelv szerinti eloszlása látható a 3. táblázatban, ahol az angol nyelvűek kiemelkedően magas részesedése tűnik fel.

Az angol nyelvű adatbázisok részarányát azonban az útmutatók kissé eltúlozzák, mert az útmutatókat angol anyanyelvű országokban készítik. Ezt a gyanút megerősíti a Dianeguide nevű adatbázis szerint elérhető európai adatbázisok nyelvi eloszlása (4. táblázat). A Dianeguide az Európai Gazdasági Közösség saját, ECHO nevű hostjánál meglévő adatbázisok útmutatója (nyomatásban is kapható). Az európai adatbázisok az EGK-országok nyilvános adathálózatain férhetők hozzá az ECHO-központban (Luxemburg). A táblázatban látható nyelvi eloszlás a Dianeguide 1989. decemberi online állapotát mutatja úgy, hogy a többnyelvű adatbázisok minden egyes nyelvnél fel vannak tüntetve. Ebből is látható az angol nyelv domináns volta, de a Dianeguide a valószínű helyzetnek inkább megfelelően tudja azonosítani az európai kontinens angoltól eltérő nyelvű adatbázisait, mint az amerikai útmutatók (vö. az 1. és a 4. táblázat eltérő adataival). A Dianeguide szerint az európai adatbázisokban az angoltól különböző nyelvű adatbázisok száma több az angolnál (persze, sok a többnyelvűségből adódó átfedés).

A bemutatott táblázatok csak az adatbázisok számát tüntetik fel, méreteiket nem. Nagy valószínűséggel a nagyobb méretű (több rekordot tartal-

2. táblázat

A nyelvi eloszlás az adatbázis típusai szerint (zárójelben a százalékok)

Nyelv	Bibliográ- fiai	Útmutató	Teljes szövegű	Az adatbázis típusa				
				Numerikus	Statisz- tikai	Grafikus	Hirdető- tábla	Tranzak- ciós
Angol	852 (78)	415 (82)	1055 (91)	769 (93)	189 (82)	3 (75)	97 (99)	27 (100)
Francia	118 (11)	34 (7)	35 (3)	15 (2)	9 (4)	1 (25)	–	–
Német	76 (7)	32 (6)	25 (2)	8 (1)	2 (1)	–	–	–
Spanyol	20 (2)	13 (3)	14 (1)	1	–	–	–	–
Japán	10 (1)	6 (1)	9 (1)	14 (2)	4 (2)	–	–	–
Egyéb	12 (1)	4 (1)	19 (2)	22 (2)	26 (11)	–	1 (1)	–
Összesen	1088	504	1157	829	230	4	98	27

3. táblázat

A CD-ROM adatbázisok nyelvi eloszlása

Nyelv	A CD-ROM adatbázisok száma
Angol	293
Német	8
Francia	8
Holland	1
Ógörög	1
Soknyelvű	12
Kétnyelvű	10
Összesen	333

4. táblázat

A Dianeguide-ban található európai adatbázisok nyelvi megoszlása\* (1989. dec.)

Nyelv	Adatbázisok
Angol	415
Német	226
Francia	215
Olasz	88
Spanyol	63
Holland	47
Dán	12
Portugál	9
Görög	4
Svéd	3
Finn	3
Norvég	2

\* A többnyelvű adatbázisok mindegyik nyelv alatt fel vannak tüntetve.

mazó) adatbázisok között nagyobb az angol nyelvűek aránya, mint a kisebbek között. 1976-ban egy 337 adatbázisra kiterjedő felmérés szerint az évi

100 000-nél több rekorddal bővített adatbázisok esetében 18 közül 15 angol nyelvű, az 50 000-nél nagyobb bővítésű adatbázisok esetén pedig 11-ből 8 az angol.

#### Az adatbázisok használata és a nyelvi nehézségek

A nyelvi nehézségek egyik nyilvánvaló következménye, hogy a felhasználó nem tudja a visszakeresett információt alkalmazni. Ez elsősorban a teljes szövegű adatbázisoknál, de a bibliográfiai adatbázisok hasznosításában is akadályt jelent. Sajnos, ennél többről is szó van. A nyomtatott dokumentumokkal ellentétben az online rendszerekben szakkifejezésekkel (egyszerű szavakkal és összetett kifejezésekkel) folyik a releváns rekordok keresése. Olyan szakkifejezéseket kell találni, amelyek a keresett információra leginkább jellemzőek, és megfelelnek a kiválasztott adatbázisnak is. Közben figyelemmel kell lenni a szinonimákra és homonimákra, az egyes és többes számra, a brit és amerikai helyesírásra, a tágabb és szűkebb értelmű kifejezésekre stb. Ezek még az anyanyelvükön dolgozó online keresőknek is nehézséget jelenthetnek, hát még az idegen nyelven dolgozó felhasználóknak!

A természetes nyelvű online keresés mellett persze mód van a szabályozott információkereső nyelvek (tárgyszavak, deskriptorok stb.) használatára is. Ilyenkor a jó indexelő rendszer részben meg tudja oldani az idegen nyelven keresők problémáit. A műszaki és természettudományokban kevesebb terminológiai probléma adódik, mint a társadalomtudományokban (gondoljunk csak az utóbbi öt évben kialakult, az egységesebb megértést segítő új szakkifejezésekre). Sokféle kódrendszer is segíti a keresést (pl. a MARC adatbázisokban, a bioszisztematikus kódok a BIOSIS-ban). A nyomtatott és online teauruszok (amelyek közül egyeseket több nyelvre is lefordítottak) sokat segítenek a releváns rekordok megtalálásában. Egyre több két- és többnyelvű adat-



bázis áll a keresők rendelkezésére. Számos angol adatbázis közli az eredeti nyelvű címet is (szavai invertáltan is kereshetők az adatbázis indexében). A PASCAL adatbázisok a francia címen kívül angol címeket is közölnek (az eredeti nyelvű cím mellett), sőt, a deskriptorok angolul, franciául vagy spanyolul kereshetők. Van már angol és francia kivonatot egyaránt tartalmazó adatbázis, és pl. a *Wer Liefert Was* nevű CD-ROM adatbázis akár angolul, akár németül kereshető.

A rekordok különböző nyelvű párhuzamos adatmezői nagy segítséget jelentenek, de egy többnyelvű adatbázis készítése elég sokba kerül. Az EGK egy másik utat választ: a kivonatok viszonylag olcsó gépi offline fordítását kívánja bevezetni. A CEC ezért megvizsgálta a SYSTRAN nevű gépi fordítórendszer hozzákapcsolását az európai hostokhoz, aminek a segítségével a felhasználó bármely bibliográfiai adatbázisban visszakeresett rekordok címeit és kivonatait az általa kijelölt nyelven kapná kézhez. 1982-ben, az első próbaüzemek alkalmával a SYSTRAN angolról franciára, franciáról angolra és angolról olaszra fordított, mégpedig az embernél gyorsabban és olcsóbban, viszont sokkal gyengébb minőségben. Sajnos, üzemszerű alkalmazására – pénzügyi és adminisztratív okok miatt – eddig nem került sor.

A CEC továbbfejlesztve gépi fordítási projektjét, további kísérleteket folytatott fordítórendszerekkel. Pl. az EUROTRA projekt jobb minőségben fordít az EGK kilenc nyelvéről és nyelvére, mint a SYSTRAN.

Az automatikus fordítás nem küszöböli ki a nyelvi nehézségeket a keresőszavak meghatározásában és a találatok online ellenőrzésében, csak az információk használatát segítené. A keresés támogatására különféle szoftvereszközök fejlesztése szükséges.

### Visszakereső szoftvertermékek

Egy parancsnyelv természetes nyelvű parancsokat használ, de nagysága erősen korlátozott. A sokféle parancsnyelv valóban okoz nehézséget a kereső személyek anyanyelvétől függetlenül is, ezt azonban nem lehet elkerülni. A parancsok megtanulása és memorizálása nem nagyon nehéz akkor sem, ha nem a kereső személy anyanyelvén hangzanak el. Ennek ellenére némely host többnyelvű parancsokat enged meg adatbázisai kereséséhez. A Dianeguide adatbázis pl. az EGK nyolc hivatalos nyelvén írt parancsokkal kereshető az ECHO-hostnál.

A CEC 1980-ban közös parancsnyelvet (*Common Command Language = CCL*) fejlesztett ki azzal a céllal, hogy az EGK hostjait ezzel használják. Azt kívánta elérni, hogy a felhasználónak csak egy parancsnyelvet kelljen elsajátítania. Sajnos, csak kevés európai host vezette be a CCL-t, sőt ezek egy része annak különféle variációit dolgozta ki.

A soknyelvű intelligens interfész (*Multilingual Intelligent Interface = MITI*) projekt célja mind az idegennyelv-ismeret hiányosságából, mind a parancsnyelvek sokféleségéből adódó online keresési problémák egy részének kiküszöbölése lenne. A

személyi számítógépre tervezett rendszerrel számos host adatbázisai válnak a gyakorlatlan kereső számára is elérhetővé azzal, hogy angol, francia, német vagy spanyol természetes nyelvű interfészt ajánl online kereséshez. A MITI-vel számos művelet hajtható végre, pl. adatbázis-kiválasztás, keresőstratégia megalkotása, bekapcsolódás és a keresés interaktív irányítása.

A menüvezérelt szoftverek – nem várt módon – feltehetőleg több nyelvi nehézséget okoznak, mint a magasabb rendű parancsnyelvek, holott céljuk éppen az lett volna, hogy egyszerűsítsék az online keresést. Sőt, minél intelligensebb egy interfész szoftver, annál több problémát okozhat annak a felhasználónak, akinek nehézségei vannak abban a nyelvben, amelyben a szoftver a felhasználóval kommunikál. Ez igaz a menüvezérelt CD-ROM adatbázisokra is.

### Szerződések, dokumentáció, oktatás

Az online keresés nyelvi nehézségei más oldalon is megmutatkoznak. Már akkor kezdődnek, amikor megszületik a döntés valamely szolgáltatóval való szerződéskötésre. A jogi dokumentumok (amilyen egy szerződés) megértése még az aláíró anyanyelvén is nehéz, hát még annak, aki számára idegen nyelven írták. Hasonló a helyzet a számlákkal, habár ezek megértése egyszerűbb (itt inkább az árkonstrukciót kell megérteni). A szolgáltatók nem nagyon törik magukat többnyelvű szerződési űrlapok kibocsátására.

Az online használati segédletek, mint a host parancsnyelv-dokumentációi, az adatbázis-leírások, a host és az adatbázis-készítők hírlevelei stb. kifejezései, nyelvezete annyira sajtóságosak és egymástól is eltérőek, hogy még ha azokat le is fordítanák más nyelvre, talán kevésbé lehetne megérteni, mint az eredeti nyelvűeket. Az írott oktatási és dokumentációs anyaggal egyenértékű a szóbeli oktatás. A tanfolyamok – akár kezdők, akár haladók számára is tartják – rövidek, tömörek, intenzívek és drágák. Az oktatás nyelvének nem tökéletes ismerete, akár az oktató, akár a hallgató részéről, nemigen vezet használható tudáshoz. A keresés közben adódó nehézségek esetén igénybe vehető telefonos tanácsadó szolgálatot (help desk) csak akkor érdemes felhívni, ha a kereső személy tökéletesen érti a hostnál jelentkező tanácsadó személy anyanyelvét, ill. fordítva; tudjuk, hogy idegen nyelven a telefonálás a legnehezebb.

Az EGK-országok 500 felhasználója körében végzett felmérés szerint a válaszolóknak jóval több mint a fele tartja úgy, hogy a szerződések, a dokumentációk és a tanfolyamok legyenek a felhasználók anyanyelvén (pl. az oktatásban 78%-uk kívánja ezt). A számlákat kb. egyharmaduk szeretné anyanyelvén kapni.

E tények ellenére az adatbázis-készítőknek és a szolgáltatóközpontoknak nemigen akarózik egy nyelviileg diverzifikált piacot kiszolgálni. Előfordul,

hogy egyes hostok hírlevelükben az angoltól eltérő nyelvű rovatokat is közölnek (pl. a Dialog host Chronolog c. hírlevelének spanyol rovata van). A felhasználói kézikönyvek azonban nem jelennek meg többnyelvű változatban (akár online, akár CD-ROM kézikönyvekről van szó).

Az online tanfolyamokat szokták helyi nyelven tartani. A Dialog host pl. 1989-ben spanyol tanfolyamot tartott Mexikóban és Barcelonában, portugált Brazíliában és Lisszabonban, dán nyelvűt Koppenhágában, németet Hamburgban és Bécsben, finn nyelvűt Helsinkiben, franciát Lille-ben és Párizsban, olaszt Milánóban, norvéget Oslóban, svédet Stockholmban (de angolt Athénben). Ennek ellenére a Dialog több angol tanfolyamot tartott Londonban, mint bármely egyéb európai városban 1989-ben. A Data-Star host telefonos tanácsadó és közönségszolgálatot üzemeltet hét európai országban, az Egyesült Államokban és Japánban, helyi nyelven.

Az, hogy a nyelvi akadályoknak ez a csoportja inkább adminisztratív, mint operatív jellegű, nem változtat azon, hogy számos felhasználó inkább lemond valamely host vagy adatbázis használatáról, mintsem hogy ilyesfajta nyelvi nehézségekkel küszködjön.

#### **Az eredeti dokumentumok nyelve a bibliográfiai adatbázisokban**

Az online kereséseknél tapasztalt nagyfokú teljesség és az adatbázisok által feldolgozott dokumentumok földrajzilag széles köre azt eredményezi, hogy a találatok sokféle nyelvű dokumentumokra hivatkoznak, amelyek egy része – az idegen nyelv miatt – a felhasználó számára értelmetlen. Ilyenkor kidobott pénz a rekordok visszakeresése, ill. a dokumentumok beszerzése, hacsak nincs jó fordítószerződés.

Az 5. táblázatban látható az angol nyelvű dokumentumok aránya néhány nagy bibliográfiai adatbázisban. Az angol nyelv részaránya a Compendexben a legmagasabb, a második helyen áll az Inspec, majd ezt követi a Biosis. Az angol után a leggyakoribb nyelvek az orosz, a német és a francia (nem ebben a sorrendben).

5. táblázat

**Néhány bibliográfiai adatbázis dokumentumbázisának nyelvi megoszlása (1987)**

Adatbázis	Témakör	Angol dokumentumok (%)
CA Search	Kémia	64
Biosis Previews	Biotudományok	85
Medline	Orvostudomány	77
Inspec	Fizika	87
Compendex	Mérnöki tudományok	91
LISA	Könyvtár és információtudomány	62
MLA	Irodalom	58
Sport Database	Sport	85

A táblázatban feltüntetett adatbázisokat angol nyelvű országokban (USA, Nagy-Britannia) készítették. De még a francia PASCAL adatbázisokban is 82%-nyi angol nyelvű dokumentumot dolgoztak fel, ezt követi a francia 8%-kal és az orosz 5%-kal. Tehát a bibliográfiai adatbázisok használói feltehetőleg túlnyomó többségben angol nyelvű dokumentumokat keresnek vissza (a maradék főleg 2–3 európai nyelven íródott). Ebből az következik, hogy hátrányban vannak, akik nem tudnak angolul, de a más nyelven nem tudók is szembe kerülnek a nyelvi akadályal (ne feledjük, hogy sok visszakeresett találat esetén a 10%-nyi más nyelvű dokumentum is sok információ elvesztését jelenti).

Néhány adatbázis hadat üzent a nyelvi nehézségeknek. A World Translation Index adatbázis azokat a tudományos és műszaki cikkeket tartalmazza, amelyeket bármely nyelvről egy nyugati nyelvre fordítottak le; ezek 70%-át angolra, 50%-át oroszról angolra fordították le. A deskriptorok angol nyelvűek. Az ECHO host Eurodicautom adatbázisa egy soknyelvű terminológiai és adatbank, amelyben tudományos és műszaki szakkifejezések, definíciók, rövidítések feloldása, nyelvi kifejezések találhatóak az EGK összes hivatalos nyelvén. Létezik néhány CD-ROM adatbázis a nyelvi nehézségek leküzdésének segítésére (pl. a Languages of the World, 12 nyelv összesen 18 szótárának teljes szövege).

#### **Következtetések**

Az online információkeresés idegen nyelv-küszöbe legerősebben Nyugat-Európában érezhető, ahol a sokféle nyelv akadályozza az információipiac egységét. A CEC erőfeszítése ellenére, hogy leküzdje ezt az akadályt, csak az angol nyelvű információszolgáltatásoknak van világszerepe. Az angol a világ népessége 8%-ának anyanyelve, 30%-a szerzett többé-kevésbé tökéletes angol nyelvtudást, vagyis az angol a legelterjedtebb második nyelv. Azok a hostok és adatbázisok, amelyek nem angol nyelvűek, nemcsak az észak-amerikai piacról zárják ki magukat, hanem a többi piac nagy részéről is. Az információközvetítő szakemberek szaktudásának jelentős részét a nyelvtudás képviseli; a nyelvészeti szaktudás a közvetítők kiválasztásának elsőrendű kritériuma. De akkor mi van a végfelhasználók ezreivel, akik a számukra idegen nyelvű adatbázisokkal és rendszerekkel birkóznak?

Az online és CD-ROM információs szolgáltatások teljes kapacitása csak akkor ismerhető fel, ha a szolgáltatóközpontok és az adatbázisok készítői maguk is több figyelmet fordítanak a nyelvi nehézségekre. A soknyelvű interfészprogramok (pl. a MITI) alkalmazása az adatbázisok keresésére részleges megoldást jelent. Az online ipar fő részeseinek ennél többet kell tenniük a soknyelvű adatbázis-hozzáférés érdekében, legalább a legfontosabb nyelvek, mint a francia, német, spanyol, orosz (sőt japán és kínai) alkalmazhatósága céljából, mind online keresésre,

mind a keresés eredményeinek értékelésére. Sajnos azonban kétséges, hogy hajlandók-e az online ipar vállalatai investálni az észak-amerikai és a nyugat-európai kulcsországokon kívüli ügyfelek érdekében.

A fenti cikkben fehéren-feketén sehol nem mondták ki, csak kerülgették a fő tanulságot. Azt, hogy aki online és CD-ROM adatbázisokban hatékonyan kíván információt keresni, és a visszakeresett dokumentumokat használni, az *tanuljon meg angolul*. Még hozzá nemcsak felületesen, hanem alaposan, hogy gyorsan tudja olvasni a kikeresett rekordokat a képernyőn, és hogy kellő változatossággal tudjon kulcsszavakat meghatározni. Ez nemcsak a közvetítőkre, hanem a végfelhasználókra is igaz, de ugyanígy igaz

a teljes K + F tevékenységre, sőt, az üzleti-gazdasági életre is. Még fokozottabban igaz ez a megállapítás a "kis nyelvek" esetén, amilyen a magyar is, amelyekre még próbálkozás sincs – mint az angoltól eltérő nagy nyelveknél – a fordítás vagy a többnyelvűség megoldására. A fenti cikk, de más hasonló tárgyú cikkek sem tesznek említést a világhozzértől eltérő anyanyelvű keresők problémáiról. – A ref.

**/LARGE, J. A.: The foreign-language barrier and electronic information. = Online Review, 14. köt. 4. sz. 1990. p. 251–266./**

(Roboz Péter)

## A Japán Tudományos és Műszaki Információs Központ angol adatbázisa

### A JICST-E adatbázis

A Japán Tudományos és Műszaki Információs Központot (Japan Information Centre of Science and Technology = JICST) 1957-ben alapították. A JICST online információs rendszere, a JOIS (JICST Online Information System) a tudományos, műszaki és orvosbiológiai szakirodalomhoz való széles körű online hozzáférést biztosítja Japánban a hazai szakirodalom feldolgozásával. A JICST online adatbázisai 1976 óta hozzáférhetők.

A JICST angol nyelvű adatbázisa, a JICST-E (E = English), a teljes japán nyelvű, tudományos, műszaki és orvosbiológiai szakirodalom 65,8%-át felölelő bibliográfiai adatbázis. 1986 óta férhető hozzá Európából, a JOIS-on vagy az STN International szolgáltatón keresztül. A cikkek angol nyelvű bibliográfiai leírásain kívül a rekordok angol nyelvű referátumot is tartalmaznak. A JICST-E forrásai: 400 folyóiratcím, kutatási jelentés és konferencia-előadás, a források közül mintegy 1000 orvosbiológiai témájú. Ezzel szemben a Medline és az Embase 200, ill. 150 japán orvosbiológiai tárgyú folyóiratot dolgoz fel. Egy tipikus rekord a JICST-E adatbázisból az *1. táblázatban* látható.

1. táblázat

#### Tipikus JICST-E rekord

AN =	Hozzáférési szám
TI =	Cikk címe angolul
AU =	Szerző(k) neve angolul és a szerző(k) munkahelye
SO =	Folyóiratcím angolul, folyóirat kód, ISSN, kötet- és füzettség, oldalszám, publikálás dátuma, táblázatok, ábrák és hivatkozások száma
CY =	Publikáló ország (Japán)
DT =	Adatok típusa, vagyis folyóiratcikk, rövid közlemény, áttekintő közlemény, előadás-kivonat
LA =	A publikálás nyelve (többnyire japán vagy angol)
AB =	A cikk kivonata
CC =	Angol teauruszból vett osztályozási kódok
CT =	Kötött tárgyszavak (deszkriptorok) angol teauruszból (legálább tíz tárgyszó)

### Az összehasonlítás módszere és eredményei

A JICST-E orvosbiológiai állományát összehasonlították az ugyancsak orvosi témájú, jól ismert Medline és Embase adatbázisokkal.

Az Embase és a Medline adatbázishoz a Data-Star-ból, a JICST-E-hez az STN-ből fértek hozzá. Hat – egy keresőkifejezésből álló – keresést végeztek a három adatbázisban. A keresések témája:

- ▶ Diclofenac (reuma elleni gyógyszer),
- ▶ Lymphokin-aktivált killersejtek (rákterápiához),
- ▶ Tumor necrosis factor (endogén protein),
- ▶ Atrial natriuretic peptid (cardiális hormon),
- ▶ Wolf–Parkinson–White-szindróma (cardiális arhythmia),
- ▶ OK-432 (immunostimuláns, klinikai próbák alatt).

Az Embase-ben és a Medline-ban csak a japán eredetű cikkek (angol vagy japán) keresésére irányult a munka, az 1985-ös és 1986-os állományra.

Az összehasonlítás eredményei a *2. táblázatban* láthatók. A három adatbázisban végzett hat keresés eredményeink összesítése szerint 346 exkluzív rekord adódott (vagyis nem átfedések), amelyek közül 272 angol cikk, 74 japán cikk. Az *1. és 2. ábra* mutatja a japán, ill. az angol nyelvű cikkek százalékos megoszlását a JICST-E adatbázisban, ill. az Embase-ben és Medline-ban együtt (E+ M).

Figyelembe kell venni, hogy a JICST-E állományának mindössze negyedrésze az orvosbiológiai szakterület, a másik két adatbázis viszont teljesen arra orientált. A JICST-E 275 cikket (127 folyóiratból), az Embase és a Medline együtt 146 cikket (52 folyóiratból) eredményezett.

### Értékelés

A vizsgálat célja az volt, hogy a JICST-E használhatóságát ellenőrizték a szakirodalom-keresésben. Már szó volt arról, hogy a JICST-E csak Japánban publikált (angol vagy japán nyelvű) közleményeket dolgoz fel, nem úgy, mint a JICST többi adatbázisa,