

## Szabadalmi adatbázisok a K+ F Információs Infrastruktúra Fejlesztési Programban

*A TMT szerkesztőségének két indítéka is van arra, hogy az olvasók különös figyelmébe ajánlja az alábbi cikket. Az egyik: átfogó képet fest a létező, a fejlesztés alatt álló és a tervezett szabadalmi adatbázisokról. A másik: a szabadalmi adatbázis-építés nézőpontjából a cikkben feltárulnak a K+ F Információs Infrastruktúra Fejlesztési Program, e várhatóan mind közkeletűbbé váló támogatási forma előnyei és hátrányai. Ez utóbbiak – remélhetően – "gyermekbetegségek", illetve a későbbiekben kiküszöbölhetők.*

### Előzmények – a PDP-korszak

Az Országos Találmányi Hivatal (OTH) "elektronizálására" vonatkozó elképzelések idestova egy évtizeddel ezelőtt merültek fel. Ezek arra a feltételezésre épültek, hogy a számítógépes rendszer egyrészt segít a hivatali ügyintézés hatékonyabbá tételében, másrészt csökkenti a hatósági munkában az átfutási időt, amelyről mellékesen megjegyezzük, hogy nemzetközi összehasonlításban jónak volt mondható. A rendszerelképzelésben természetesen elsőbbséget élvezett a publikus, "közhasznú" adatbázisok létrehozása. Az OTH-nak ekkor nem volt saját számítógépe. Ügyfélforgalma és költségvetése nem is kecsagetett azzal, hogy valamennyire is számottevő saját számítógép-kapacitást építhessen ki.

Ez az időszak, a 80-as évek eleje, Magyarországon a hierarchikus, centralizált nagygépes rendszerek telepítésének, hasznosításának korszaka volt. Több cég egymással versengve ajánlotta közhasználatra nagygépeinek kapacitását. Az ezen cégek szervezői által készített rendszertervek általában olyan megoldásokat kínáltak, hogy a "távoli" nagy számítógépre az OTH adatátviteli vonalon terminálokkal csatlakozzék.

Az adatállomány gyarapodásának ütemét felbecsülve azonban nyilvánvalóvá vált, hogy sem az induláskor, sem az utána következő években nem lesz szükség a nagyszámítógépes erőforrások által kínált nagy (és drága) központi memória- és háttértárkapacitásokra. Ennek megfelelően az ügyviteli rendszer helyi központjaként az OTH épületében elhelyezett, a KFKI által gyártott, PDP-11 kompatibilis TPA 11/48 típusú számítógépre esett a választás, amely az Alkotó Ifjúság Egyesülés tulajdonában volt. Az OTH VT-52 terminálokkal kapcsolódott a gépre és nagy felhasználóinak egyike volt.

Az OTH döntéshozói és a rendszer fejlesztői rövidesen érezkelhették, hogy e megoldással ingoványos talajra léptek. Akkortájt ugyanis rendkívül nehezen megoldható problémákat jelentett a magyar karakterkészlet korrekt kezelése (képernyős megjelenítése, kinyomtatása, indexbe sorolása). Az úttörők

gondja a rendszer optimális szervezése, az adattárolási konvenciók kialakítása tekintetében sem volt kisebb, ami nem is tekinthető olyan triviálisnak, ha például az elsőbbségi aktaszámok szinte hivatalonként eltérő írásmódjára gondolunk.

### A mikroszámítógépes korszak felé

Mai szemmel nézve némileg furcsa az a 80-as évek közepén tett javaslat, amely egy jó nevű számítástechnikai vállalat szervezőitől származott, miszerint a VT-52 terminálok és a TPA online kapcsolata helyett adatbeviteli célokra fontolóra kellene venni a C64-es személyi számítógépek alkalmazását, miközben csak a lokálisan rögzített adatok "batch" (köteget) feldolgozása "maradna" a TPA-ra. A ma már megmosolyogható javaslat valójában az akkori országos helyzet kritikája. Azon döntéshozók felelősségére emlékeztet, akik a házi és játékszámítógépek forgalmazására gyorsan rámozduló cégek számára olyan extraprofitot ajándékoztak, amelyet a hazai fejlesztési kényszer alatt álló vállalatokkal és intézményekkel fizettettek meg. Az OTH ebbe az utcába végül is nem ment be.

Ekkorra – a TPA-s ügyviteli rendszer alkalmazásának első tapasztalatai birtokában – már látni lehetett a korábbi fejlesztések erényei mellett a gyengéit is, nevezetesen, hogy az ügyviteli rendszer nem tudta leképezni a teljes ügykezelési folyamatot, és hogy az elektronikus ügykezelés csak a szabadalmi bejelentésekre terjedt ki. A védjegy és ipari minta bejelentésekkel kapcsolatos ügyforgalom mérete ezt a diszkriminációt még minden továbbiak nélkül elviselte volna, ha nem bizonyosodott volna be, hogy az OTH e tekintetben egy, a századelőre visszanyúló ügykezelési tradíciót és kultúrát testesít meg, amelynek egy még nem teljesen érett technikával való felváltása a jövő igényesebb feldolgozási követelményei szempontjából súlyos kockázatokat jelenthet. A számítógépesítés a hivatali tevékenységnek csupán egy – igazából persze nagyon fontos területét – tudta csak lefedni: a bejelentőkkel (és felalálókkal) való levelezést.

Az persze, hogy számítógépesített ügyviteli rendszerünk levelezési típusú rendszer lett, hogy az OTH egy csapásra nem vált a világ elektronizált hivatalainak egyikévé, még nem lett volna nagy gond. Annál fájdalmasabb volt viszont az az időnk folyamán egyre világosabbá váló felismerés, hogy miközben az ügyviteli rendszer tartalmazza a "közhazsnú" tájékoztatás majdnem minden fontos ismeretét, valójában sem az adatok tárolási formátuma, sem az alkalmazott adatbázis-kezelő rendszer, sem a hivatali ügyrend nem alkalmas egy publikus nemzeti szabadalmi adatbázis illetően létrehozására.

## Levelezési ügyviteli PC-rendszer: OTHello

Az áttörést a levelezési rendszer fejlesztésében a professzionális személyi számítógépek (PC) korszaka hozta. Az "alapadatok" 1987–1988-ban a TPA gépről átkerültek egy IBM személyi számítógépekből álló helyi hálózatra (LAN). A hálózat középpontjában két 40–40 Mbájtos winchesterrel felszerelt gép állt, melyre tucatnyi PC kapcsolódott – intelligens terminálként. A TPA ettől kezdve már csak a tradíciók miatt látott el olyanféle rutinfeladatokat, amelyeket a PC-LAN is el tudott volna látni.

A személyi számítógépek hálózatára telepített ügyviteli-levelezési rendszer neve OTHello lett. Az OTHello működtetését az akkor szinte mindenütt alkalmazott standard relációs adatbázis-kezelő rendszerre bízták, amelynek már a kezdetekben is látott gyengéje, hogy szöveges adatok kezelésére (például szavas indexelés alapján való visszakeresésre) nem alkalmas.

További probléma volt a rendszerszervezés gyengesége. A levelezés szempontjából korrekt megoldások az adattárolási konvenciókat illetően a közhazsnú adatbázisok számára már távolról sem voltak megfelelők. Ez akkor vált világossá, amikor a Szabadalmi Közlöny és Védjegyértékesítő szabadalmi részét, pontosabban annak "Adatközlés szabadalmi bejelentésekről" rovatát az OTHello kezdte el adatokkal táplálni. A számítógép által előállított lista ekkor még csak a kéziratot helyettesítette: a nevek benne sokszor csupa nagybetűvel voltak írva, a címek furcsán "elharapott", be nem fejezett formában jelentek meg, az alkalmazott kódkészlet csak a szűkített magyar ábécét tudta produkálni, a névállomány nem volt egységes, ugyanaz a név és lakcím hol így, hol úgy szerepelt, az elsőbbségi aktaszámok tárolási formátuma a bejelentő által adott adat betűhív megjelenítése volt stb.

E problémák egy része korrigálható volt. Sikerült például megoldást találni a hosszú nevek és találmánycímek tárolására, a Nemzetközi Szabadalmi Osztályozás (NSZO) jelzeteinek helyi értékes kezelésére. Van azonban egy olyan része a bajoknak, amelyekkel amíg az OTHello él, együtt kell élnünk. Ide tartozik például a névállomány divergenciájának problémája. Minthogy az OTH – tudatosan és dicsé-

retes módon – a bejelentő, feltaláló személyi számát nem tárolja, a névállomány pedig a levelezés lelke, kívánatos volt, hogy a név- és lakcímadatok minden bejelentéskor a megadott betűhív formában kerüljenek az adatbázisba. Mindez a levelezési rendszer céljaira tökéletesen megfelelt, de a közhazsnú rendszer létrehozásában szinte leküzdhetetlen akadályt jelentett. Hasonló gondot okozott a transliterálási szabályok korai időkre jellemző be nem tartása, alkalmazásának esetlegessége. Ilyen probléma nyilvánult meg az elsőbbségi adatok tárolásában is, ami a szabadalomtisztasági vizsgálatok ellehetetlenítésével volt egyenértékű.

## A HUNPADOC megszületése

A Magyar Iparjogvédelmi Egyesület egyik 1985-ben, az MTA Számítástechnikai és Automatizálási Kutatóintézetében tartott rendezvényén az OTH bemutatta TPA-alapú ügyviteli rendszerét. A bemutató után a Számítógépes Kémiai Informatikai Társaság (SZKIT) egyik munkatársa felhívta a jelenlévők figyelmét a magyar szabadalmi bejelentések általuk forgalmazott adatbázisára, amelyet a CompuDrug által fejlesztett LIBASE nevű program kezel, és egy személyi számítógépen a kémiai szakterület összes Magyarországon tett bejelentésének adatait tartalmazza.

A rendszer alapadatai a Kőbányai Gyógyszerárugyárban korábban ODRA gépen feldolgozott állomány konverziójából származtak. Az adatbázis adatellátását a Szabadalmi Közlöny alapján végzett utólagos adatrögzítés biztosította. Nyilvánvalóvá vált, hogy amit az OTH több év alatt saját rendszeréből egyáltalán fel tudott volna építeni, az igen jelentős archív állománnyal máris rendelkezésre áll. A SZKIT-adatbázis valójában egy nemzeti szabadalmi adatbázis alapjaként kínálkozott, csak hogy az akkori PC-árrakkal (akadt pl. 1 millió Ft-nál drágább PC is) összevethető, elkedvetlenítően magas áron. Ez, továbbá az a tény, hogy az adatbázis lényegében egy PC-t teljesen lekötött, rendkívül nehezé tette forgalmazását.

1987-ben az OTH-SZIK megállapodott a CompuDrugból INFOKER Kiszövetkezett néven kivált fejlesztőkkel a SZKIT-adatbázis átdolgozásának finanszírozásáról. A megállapodás példászerű volt. Az adatállomány konverziója magában foglalta a csupa nagybetűvel, ékezetek nélkül rögzített címek ékezetesítését is, ami nem is annyira pragmatikus követelmény volt, mint inkább az anyanyelv védelmének és tiszteletének a követelménye.

Az átdolgozással párhuzamosan a CompuDrug/INFOKER teljessé tette az adatállományt, amely így már 1970-ig visszamenőleg az összes magyar szabadalmi bejelentés, megadott és megszünt szabadalom adatait tartalmazta. Ezenfelül a feltalálók korábban nem rögzített névadatait is felvették az adatbázisba.

Az így létrejött adatbázist a továbbiakban az OTH és az INFOKER közösen kezdte forgalmazni. Az adat-

bázis, amely az INFOKER saját fejlesztésű FULLTEXT adatbázis-kezelőjére épült, már számot tarthatott arra, hogy igazi nemzeti szabadalmi adatbázisnak tekintessék. Ezt jelzi új neve, a HUNPADOC. Ez a név emlékeztet arra, hogy az adatbázis a magyar bejelentésekben és szabadalmakban szereplő adatokat a világszerte közismert International Patent Documentation Centre (INPADOC) által használt adattárolási mélységben tartalmazza. (Egyetlen lényeges korlátja, hogy szöveges kivonatok nincsenek benne, ahogy az INPADOC-ban sincsenek.)

A HUNPADOC egyik legfontosabb erénye, hogy az adatközlés stádiumában lévő ügyek számítógépes keresését is lehetővé teszi, illetve hogy a szabadalmak jogállására vonatkozóan is tartalmaz státuszjelzést. Nem hagyható említés nélkül a beépített statisztikai elemzés parancsa sem. Ez a nagy adatbankok statisztikai elemzési parancsainak és a Derwent Publications Ltd. PATSTAT programcsomagjának fejlesztéseit vette mintául. A statisztikai elemzés segítségével a vállalati felderítés olyan tájékoztató eszköz birtokába jut, amelynek segítségével a korábban alig megválaszolható kérdésekre szinte gyerekjáték lett választ adni.

### A K+ F Információs Infrastruktúra Fejlesztési Program

Az Országos Műszaki Fejlesztési Bizottság és a Magyar Tudományos Akadémia támogatásával 1988-ban indul útjára a K+ F Információs Infrastruktúra Fejlesztési Program. Fő célkitűzése a kutatókat egymással és a program keretében kiépítendő adatbázisokkal, valamint a külföldi adathálózatokkal, üzenetkezelő és adatátviteli rendszerekkel összekapcsolható hálózat létrehozása. Az igen gyakran I<sup>2</sup>F-ként emlegetett program keretében született meg a "Kernel rendszer", amely egy 9600 bit/s névleges adatátviteli sebességű csomagkapcsolt hálózatból, annak végberendezéseiből, valamint a szükséges hardverből áll.

A hálózat önmagában természetesen csak "lehetőség". Értéke az új infrastruktúra mögött álló adatbázisok értékéből adódik, különösen ha kiépülnek azok az elektronikus üzenetközvetítő rendszerek is, amelyek a kutatók közötti gyors és hatékony kommunikáció eszközei lesznek.

A program ideáltípusa egy olyan hálózat, amely biztosítja, hogy a különféle személyi számítógépek és terminálok előtt ülő kutatók anélkül jussanak – pl. a referálólapoknak megfelelő bibliográfiai adatbázisokból – számukra fontos szakmai ismeretekhez, illetve anélkül válthassanak üzenetet társaikkal, hogy akár egy másik intézménybe, akár saját intézetük könyvtárába ki/be/le/fel/át kellene menniük.

### Szabadalmi adatbázis pályázatok

Az I<sup>2</sup>F program nem akármilyen feladatot vállalt magára, ha meggondoljuk, hogy az Egyesült Államok nagy adatbázisai részben hatalmas nemzeti programok kiszolgálására és támogatására indított programokból nőttek ki. Közismert például, hogy a Dialog online szolgáltatóközpont az Apollo program infrastruktúráját megteremtő információs rendszerből született. Másfelől az sem lényegtelen előny, hogy a 60-as évek adatbázisainak többsége referálólapok indexeinek készítéséből vagy a referáló folyóiratok számítógéppel támogatott szedéséből, tulajdonképpen "melléktermékként" jöhetett létre, sőt jön létre mindmáig.

Magyarországon a holdraszálláshoz hasonló léptékű programról ugyan nem lehetett szó, de az I<sup>2</sup>F program keretében jelentős, mintegy 80 millió forintot fordítottak kifejezetten adatbázisok építésére. A magyar referáló folyóiratok többségét nem elektronikus szedéssel állítják elő, ha netán mégis, a lemezeket, szalagokat (takarékoságból) gondosan letörölték, tehát kész adatbázisok alig-alig vannak. Minthogy a K+ F információk között a szabadalmi információ szerepe nagyon fontos, aligha lehet meglepő, hogy a szabadalmi adatbázisok építésére is jelentős összeget irányoztak elő, összességében mintegy 10 millió forintot. E fejlesztési támogatások elnyerésére 3 vállalkozó pályázott: az MTA SZTAKI, az Ipari Informatikai Központ (INFORMATIK) és az OTH.

Az INFORMATIK egy online nemzetközi szabadalmi témafigyelő szolgálat tervével állt elő (ez a korábban ASZBA néven ismert, INPADOC mágnesszalagokra épülő offline témafigyelés folytatása volt). Az OTH a magyar szabadalmi adatbázis adatközlési, az MTA SZTAKI a közzétételi, az INFORMATIK a megadott szabadalmak szegmensének építésére pályázott.

Hosszas viták után az OTH személyi számítógépes szolgáltatásra nyert jelképes támogatást (500 ezer forintot). A fennmaradó jelentős összeg az INFORMATIK és az MTA SZTAKI között oszlott meg. A komoly számítógépes kapacitásokkal diszponáló INFORMATIK az ASZBA révén e területen "előélettel rendelkező" céggé nyert támogatást. Az MTA SZTAKI-t mindenekelőtt az hozta kedvező pozícióba, hogy az I<sup>2</sup>F program talán leginkább profiltiszta adatbázisainak forgalmazását tervezte (a KUFET, azaz a hazai kutatás-fejlesztési témák nemzeti adatbázisának fontossága vitathatatlan). Korábban a SZTAKI gondozta a program indításakor már körvonalazódni látszó újítási adatbázis, a Szellemi Termék-adatbázis fejlesztését is. Mindezekon túl a leglényegesebb szempont az volt, hogy a SZTAKI saját számítógépes nyelvészeti kutatásainak adaptálását és integrálását is felajánlotta egy igazi, a nemzetközi szabványokhoz alkalmazkodó, szöveges kivonatok is tartalmazó szabadalmi adatbázis létrehozásához.



Noha, mint láttuk, a két nagynevű professzionális számítástechnikai cég jelentős dotációs sikerrel pályázott a szabadalmi adatbázisok építésére, egy percig sem volt kétséges, hogy az I<sup>2</sup>F szabadalmi adatbázis-építési programjának gesztora az OTH kell legyen. Nagy súllyal esett latba, hogy az OTH tiszteletbeli tagként, "megfigyelő" státusban jelen van az OLPI-ban. Az OLPI az Online Szabadalmi Adatbázis-építők és Felhasználók társasága, ahol például a Bayer a felhasználók, a Derwent és az INPADOC az adatbázis-előállítók, a Dialog pedig a szolgáltatók képviselője.

Minthogy az online ipar – az üzleti érdekek sokszínűsége miatt – fittyet hányva a felhasználók ebbéli sirámaira, nem nagyon kíván kívülről szabványosítást elfogadni, az OLPI e téren irányutató kezdeményezésnek tekinthető. Mint ilyen, a korszerű technikák alkalmazásától kezdve az indexezők jelzeteinek definiálásáig sok mindennel foglalkozik. Ezeknek az ismereteknek a kamatoztatása, az itt szerzett tapasztalatokra épülő javaslatok elfogadása az egész I<sup>2</sup>F program hasznára válhat.

## Gyermekbetegségek

Az I<sup>2</sup>F keretében épült szabadalmi adatbázisok bemutatása előtt célszerű áttekinteni a projekt "gyermekbetegségeit". Az I<sup>2</sup>F első ütemében még nem lehetett világosan látni, hogy mit vállalnak az adat- és adatbázis-tulajdonosok, mit a majdani szolgáltatók képviselői, illetve mit a fejlesztők, akik az üzemeltetéshez szükséges szoftvereszközöket biztosítják. A hazai információs kultúrát sem sikerült integrálni a programba, holott az a külföldi online rendszerek intenzív használata folytán számos magyar intézménynél vagy vállalatnál rendelkezésre állt. A résztvevők között volt olyan intézmény, amely nála meglévő, "kapun belüli" ismereteket, sokszor alapismereteket hagyott figyelmen kívül. A majdani szolgáltatókat nem zavarta hároméves szerepvállalásukban, hogy a még létre sem hozott adatbázisok potenciális piaca valójában nem az I<sup>2</sup>F hálózatba kapcsolt terminálok számából, hanem a hálózatot elérő intézetek számából kalkulálható. Sokan alábecsülték a távbeszélő-hálózatról a csomagkapcsolt hálózatba való belépés lehetőségét, bár a piac fokozatos meghódításában a "kis vevőknek" legalább akkora szerepük lehet, mint az I<sup>2</sup>F fejlesztések által preferált kutatóhelyeknek.

## OLT – online találmányi adatbázis, a szabadalmi közlések adatbázisa

Az OnLine Találmányi (OLT) adatbázis furcsa, a projektből kissé kilóg. Alapja az OTH ügyviteli-levelezési rendszere, az OTHello. Az OTHello adatainak publikus része nyilván közérdeklődésre tarthat számot, hiszen közel 10 ezerre tehető a közzététel

előtt álló azon ügyek száma, amelyek más forrásból, mint a Szabadalmi Közlöny "Adatközlés szabadalmi bejelentésekről" rovata, nem ismerhetők meg. Köztudott, hogy ez a rovat a bejelentéseket NSZOSzekciónként, azokon belül a bejelentők nevének beütrendjében publikálja. Ha több bejelentő közös bejelentéséről van szó, az alapadatok csak az első bejelentő nevével jelennek meg: az ügyiratszám, a bejelentés dátuma, a bejelentő(k) neve és címe, a találmányi bejelentés címe, a képviselő neve és címe, valamint az elsőbbségi adatok (dátum, ország, akatszám). Minthogy a Közlöny e rovatának retrospektív keresése a közzététel előtti ügyek megtalálása végett másfél év állományának lineáris, "mutató által nem támogatott keresését", azaz az összes szóba jövő szekciók végigolvasását követelné meg, a legtöbben lemondanak arról az aligha túlbecsülhető előnyről, amelyhez az ilyenfajta korai szignálinformációk révén lehet hozzájutni. Ez a fajta adatközlés a nemzetközi szabadalmi jogban ritka. Értéke önmagáért beszél a következő példa tükrében: ha egy magyar elsőbbségű bejelentésre hivatkozva az egyéves uniós elsőbbség igénybevételével az Amerikai Egyesült Államokban is bejelentést tesznek, az USA-ban a magyar bejelentést követő 3 év, itthon 3 hónap múltán jelenik meg az első adat a bejelentésről. Ez pedig akkora versenyelőny, amit bűn lenne kihasználatlanul hagyni. E publikálás a "vállalati felderítés" lehetőségeit is kibővíti. Noha a bejelentést csak az elsőbbség dátumától számított 18 hónap múltán teszik közzé, tartalma tehát mindaddig nem ismerhető meg, azért a cím szintjén közzétett ismeretek előzetes tájékoztatást adnak a konkurenciaharcban állók számára. Másrészt az adatközlések – s ez bevallott céljuk – a hasznosításra irányuló kapcsolatok mielőbbi felvételét is segítik.

Az OTHello az 1984. március utáni szabadalmi bejelentések adatait tartalmazza. Felmerült a kérdés, hogy vajon csupán az adatközlési stádiumban lévő adatokat célszerű-e benntartani az OLT adatbázisban. Végül olyan döntés született, hogy az OTHello adatait – a HUNPADOC alapján – ki kell egészíteni a közzétett bejelentésekre vonatkozó adatokkal (közzététel dátuma, közzétételi szám), illetve ha a hivatal szabadalmat engedélyezett, a lajstromszámmal.

A döntésben az is szerepet játszott, hogy az alkalmazott adatbázis-kezelő rendszer ennek az archív állománynak az online kereséséhez remek eszközöket kínált. Az OTH az OLT adatbázis szolgáltatását PC-alapú szolgáltatásként képzelte el. A HUNPADOC-ot előállító INFOKER tapasztalatai meggyőzően igazolták, hogy a PC-kben rejlő lehetőségek kiaknázásával lehetséges akár több tízezer dokumentum adatait is személyi számítógépes adatbázisban hatékonyan kezelni (a HUNPADOC 70 ezer találmány adatait tartalmazza). Az INFOKER kínálta szoftvereszköz alkalmazásával a korlátozott keresési lehetőségeket kínáló ügyviteli adatbázis mérete 40 Mbájtról a harmadára, 13 Mbájtra csökkent (az adattartalom lényegtelen mértékű szűkítése – kimaradtak az

ügyintézésre vonatkozó kódok – , illetve részleges bővítése mellett – adatmezőket importáltunk a HUNPADOC-ból – , miközben a keresési lehetőségek összehasonlíthatatlanul jobbak lettek).

Az OLT adatbázist, amelyben jelenleg mintegy 25 ezer ügy bibliográfiai adatai és több mint 30 ezer feltaláló, illetve bejelentő név- és lacímadatai található meg, az eredeti elképzelés szerint egy Novell-alapú helyi hálózat termináljain belépve lehetett volna majd elérni. Ennél nem a többfelhasználós üzemmód okozott volna problémát, hanem a megfelelő adatátviteli út kialakítása. Kezdetben az OTH az X25-ös állomás alkalmazását preferálta, ez ugyanis a párhuzamos használat lehetőségét kínálta. Ez ellen szólt viszont az, hogy a hivatal számára a RADAUS felé történő kijárási, a külföldi adatbázisok online elérése nem mellőzhető szempont volt. Az alaphálózat bővítése csak 1989 júliusára tette lehetővé egy X20-as állomás telepítését hivatalunkba.

## TEXTAR

Az OLT egyik erőssége az adatbázis-kezelő rendszer hatékonysága, érdemes tehát néhány szót szólni a (TELE)TEXTAR adatbázis-kezelőről.

A TEXTAR az INFOKER Kiszszövetkezet által kifejlesztett FullText továbbfejlesztésével készült, a piacon jelenleg sok tekintetben egyedülálló program, a szöveges adatbázisok többállományos adatbázis-kezelője. A szöveges adatbázis-kezelés minden kritériumának megfelel (szavas indexelés, változó hosszúságú rekordok stb.). Alkalmazása helytakarékos, ráadásul a keresési lehetőségeket is jöcskán kibővíti. A TEXTAR több állományba sorolt adatokat kezel. Ezeket akár külön katalógusoknak, külön adatbázisoknak is tekinthetnénk. Ha például már van a könyvtárunkban folyóirat-katalógus, az analitikus cikk-katalógusba a folyóirat adatait külön nem kell bevinni, minthogy a cikk-katalógus használhatja a folyóirat-katalógus adatait. Nagyszerű keresési lehetőségeket kínál a rekordok közötti kapcsolatok kezelése. A több állományba lerakott rekordok közül bármely állomány bármely rekordja minősített kapcsolatba hozható egy másik, bárhol szereplő rekorddal.

Az OLT az OTH hatósági tevékenysége számára is értékes (például az igényrontó bejelentések megkeresésében kaphat szerepet), de – és ez sokkal fontosabb – elsősorban a köz számára nyújt jó hozzáférést a szabadalmi adatközlések kezeléséhez. A bázis adatminőségét az adatok relációs szervezése biztosíthatja. (Az OLT "szűrésének" újabb fordulója van folyamatban; lezárta a levelezési-ügyviteli rendszerrel jobb adatminőségű és konzisztenciájú rendszer áll fel.) Az INFOKER által kifejlesztett TEXTAR az adatbázis kompakt tárolását és hatékony kezelését egyaránt támogatja.

A proximity search, vagyis a szótávolságra való keresés lehetőségét leszámítva, a felhasználók a TELE-TEXTAR segítségével számos új keresőnyelvi

elemet érvényesítő eszközként vehetik igénybe az adatbázist, mely elsősorban a vállalati felderítés fegyvertárát bővíti. Segítségével könnyen kideríthető, hogy kik valamely vállalat vezető feltalálói, hogy egy kutatóintézet milyen gyakran kényszerül közös kutatásokra, hogy egy potenciális konkurens milyen iparjogvédelmi stratégiát követ. Kényelmesen figyelhető azon partnerek tevékenysége is, akik rászolgáltak a bizalmatlanságra, megtudható, hogy egy-egy régió milyen területeken érdeklődik a magyar piac iránt stb.

## SZKIV – a közzétett bejelentések kivonatokat tartalmazó adatbázisa

A SZKIV (az MTA SZTAKI-ban: SZIR = Szabadalmi Információs Rendszer) a közzétett bejelentések adatbázisa. Köztudott, hogy a Szabadalmi Közlöny a bejelentések közzétételekor adja meg a találmányi megoldást röviden ismertető szöveges kivonatot (a megadáskor megint csak a bibliográfiai adatok jelennek meg).

Az MTA SZTAKI központi IBM gépére építve, az azon installált CDS/ISIS 4.6-ra alapozva vállalkozott a SZKIV adatbázis létrehozására. Az e tárgyban létrejött szerződés megkötésekor az OTH ezt az adatbázist az egységes (az összes fontos bibliográfiai adatot egy állományban tároló) magyar nemzeti szabadalmi adatbázis modelljének tekintette. Nagy remények éltek azzal kapcsolatban, hogy az MTA SZTAKI számítógépes nyelvészeti kutatásainak eredményeit reprezentáló "SZÓTÁR" integrálható lesz a SZKIV keresőrendszerébe.

A I<sup>2</sup>F magyar szabadalmi adatbázisok építését célzó fejlesztési forrásainak fele a SZKIV létrehozásának költségeit fedezte. A modell-adatbázis egyetlen év, 1988 adataival jött létre, ami enyhén szólva, korlátozott értékűvé tette az adatbázist. Az OLT és a SZKIV között az a legalapvetőbb különbség, hogy az utóbbi nem tartalmazza a megadott szabadalmak lajstromszámait, tartalmaz viszont egy informatív kivonatot, amit az ISIS szótávolságra való keresést is támogató operátoraival nagyon jól kereshető tesz. A rögzített kivonatok alapján a SZTAKI által felépített SZÓTÁR bővítése nem történt meg, részben a szótár ékezetnélkülisége folytán, részben mert a feldolgozott állomány kicsi volt.

Ilyen kisméretű adatállománnyal az adatbázis sikerére nem lehet számítani. Éppen ezért az MTA SZTAKI vállalta, hogy az adatbázis állományát legalább 5 évre bővíti ki. Ehhez a bibliográfiai adatokat az OLT adatbázis biztosítja.

A kivonatok kezelése egyébként egy sajátos problémára világított rá. Az archív adatok előállítására visszamenőleges adatrögzítéssel történt. Rövidesen kiderült, hogy a kivonatok kereshetővé tételekor meg kell birkóznia a tipográfiai szabályok sokaságával, pl. a szövegbe tördelt képletek kifejtésekor, az alsó és felső indexek jelölésekor. Az OTH fejlesztői számára

nyilvánvalóvá vált, hogy a továbbiakban az ilyen konstrukció az adatbázis folyamatos adatellátására nem megfelelő. E zsákutca kikerülését biztosítja, hogy 1990. január óta a Szabadalmi Közlöny és Védjegyértesítő előállítására már a  $\text{MaT}_{\text{E}}\text{X}$  szedő-tördelő program (ejtsd: tekhh) segítségével történik. Az adatbázis bibliográfiai adatainak kezelését egy hivatali adatbázis látja el, melybe a nyomdában szedett kivonatok számítógéppel olvasható szövegét utólag építjük be, miközben megtartjuk bennük a  $\text{T}_{\text{E}}\text{X}$  szedési utasításokat.

A SZKIV fejlesztésében a gesztor pirruszi győzelmeket aratott. A SZKIV azon kevés adatbázisok egyike, amelyben többszintes indexelés is van (azaz a CO2B alosztályba tartozó bejelentések nem csupán a CO2B5 csonkolt kifejezéssel kaphatók meg). Sikerült elfogadtatni bizonyos adatmezők, például az elsőbbségi ország adatmező prefixes jelölését, elismertetvén az online keresők számára nyilvánvaló szabályt, hogy például a külföldi bejelentő aktivitásának becslésére más módszer nemigen létezik. (A korrektség kedvéért megjegyezzük, hogy az MTA SZTAKI időközben kifejlesztette azt a – prefix helyett – postfix kvalifikálással működő EXPAND – az ISIS-ben EXPLORE – funkciót, amely összehasonlíthatatlanul lassabban, de szintén megválaszolhatóvá teszi ezt a kérdést. Működik tehát az "EXPLORE A\$/31)" formátumú (postfixes) parancs, de az előző példánál maradva, a villámgyorsan választ adó "EXPLORE PC" = (prefixes) parancs helyett a keresés eredménye túlságosan hosszú idő múltán jelenik meg.)

## ASZBA – INPADOC-témafigyelés INFORMATIK-változatban

Az INFORMATIK jogelődje, a KG Informatik 1982-ben indította meg az ún. ASZBA-SZOBI szalagok offline feldolgozására épülő témafigyelési rendszerét, amely lényegében egy INPADOC-alapú számítógépes témafigyelés. Az I<sup>2</sup>F első ütemében az INFORMATIK e szolgáltatás online változatának létrehozásával pályázott. (A szabadalmi adatbázisok építésére szánt teljes összeg fele-harmada fedezte e projekt költségeit.) INPADOC-monitor szolgáltatások a világ nagy szolgáltatóközpontjaiban is vannak, így e szolgáltatásnak kemény versenytársai akadnak. Ráadásul a Dialog, Orbit, STN INPADOC-alapú témafigyelő szolgáltatásának egyik közvetítője éppen az INFORMATIK.

A projekt első változatában egy viszonylag szűkebb országhatár figyelésének igénye merült fel. Ez mára úgy módosult, hogy – ezzel sokan egyetértenek, mások nem – a japán adatok a továbbiakban nem szerepelnek az adatbázisban. (A kollégákat, akik a japán átírást megérteni képesek, egy kezünkön meg tudjuk számolni.) 1990-re újra változik a helyzet, a szalagokon ugyanis megjelennek a japán bejelentések angol nyelvű címei is.

Az ASZBA árai 1982-től 1989-ig, több mint fél évtizeden keresztül nem változtak. Az alacsony ár nem túl feszes szolgáltatási rend mellett is kifizetődővé tette az előfizetést, a szolgáltatás rentabilitásának azonban nem kedvezett.

Az I<sup>2</sup>F-ben a NSZ (Nemzetközi Szabadalmi témafigyelés) néven szereplő online adatbázis létrehozását heves viták előzték meg (a résztvevők álláspontja e tekintetben mindmáig különbözik), mindazonáltal az NSZ frekvenciált adatbázis lehet (ügyfeleinek száma megközelíthetné a százat). Ezenkívül az online kapcsolat az X25 hálózaton számítógéppel olvasható formában kezelhető eredményeket szolgáltat (downloading), aminek előnyeit nem kell külön ecsetelni. Az NSZ egyébként a lehetőségek figyelembevételével korrekt CDS/ISIS alkalmazás. Benne a többszintes (pl. az NSZO mezőre való) indexelés és a prefixes (pl. a bejelentők névadataiban való tallózáshoz szükséges) indexelés egyaránt megtalálható. Az offline szolgáltatásnak jól definiált ára van: 3000 Ft/év előfizetési díj mellett 10 keresőkulcsot tartalmazó profil futtatható, 9000 Ft/év előfizetési díj ("bérlet váltása") esetén korlátlan számú és bonyolultságú profil. A találati díj 13 Ft/nyomatott tétel.

## LAJSTROM – az érvényes szabadalmak adatbázisa

A szerencsétlen módon először MSZ-nek (Magyar Szabadalom) keresztelt adatbázis furcsa kettősséget mutat. Egyfelől az INFORMATIK vállalkozásában létrejött adatbázishoz a gazdaság résztvevőinek elemi érdeke fűződik, hiszen a szabadalmi jogokról való gyors tájékozódás több mint parancs. (Jelenleg mintegy 20 ezer találmányt véd magyar szabadalom.) Másfelől ezt az adatbázist valóban "közhasznúvá" tenni csak jelentős munkával lehet.

A LAJSTROM egyébként az I<sup>2</sup>F viszonylag kisebb költségű projektje volt. Adatállománya az átadásakor és jelenleg is mindössze egy év adatait tartalmazza. Egyelőre "hiányzik belőle" a Szabadalmi Lajstrom igazi számítógépes feldolgozása. A SZIK a vele kapcsolatban felismert igényeket jogosnak tartja, és 1990-ben prioritást ad a fejlesztésének. Adatokkal való ellátásában 1990-től, a Szabadalmi Közlöny "Vegyes szabadalmi közlemények" rovatának számítógépes szedése óta nincs akadály. (Az előrelépésre további alapot szolgáltat az INPADOC-kal kötött közvetlen adatellátási megállapodás, melynek értelmében hivatalunk belép a PRS adatszolgáltatók sorába.)

## Összefoglalásként

A folyamatos adatellátáshoz egy lassan bővülő piacon, egy most formálódó infrastruktúrában kell erőforrásokat és forrásokat teremteni. Az I<sup>2</sup>F adatbázisok helyzetének áttekintése után megállapítható, hogy a projektben számos gyermekbetegség lépett



fel. Ezek hatása bizonyos fejlesztési területeken erőteljesebben érzékelhető, másutt kevésbé. Sajnálatos, hogy a fejlesztés éppen abban az időszakban kezdődött, amikor a minisztériumi és más háttérintézmények maguk is komoly gondokkal küszködtek. Az I<sup>2</sup>F emiatt sok tekintetben az intézményfinanszírozás jegyeit viseli magán. A gesztoroknak, akiknek az adatbázisokkal kapcsolatosan komoly erkölcsi, mi több "ellátási" felelősségük is van, viszonylag kicsi a mozgásterük.

## A problémák megoldása: MAT<sub>E</sub>X

Van-e kitörési pont? Megítélésünk szerint igen, és ez nem más, mint az "electronic capturing", azaz az adatok megfogása a keletkezés helyén, amire az OTH közlönyének elektronikus előállítás ad lehetőséget. A Szabadalmi Közlöny és Védjegyértesítő olvasói – legalábbis azok, akik Svájc, Finnország, Hollandia és még jó néhány előtünk járó ország közlönyeit olvassák – saját szemükkel is meggyőződhetnek róla, hogy az OTH közlönyének előállításában különös szakmai műgond tükröződik. Ennek jó példája a WIPO-szabványokra való támaszkodás a rovatokra tagolásban, az adatok kiemelésében, az előfejek kezelése, az adatok tagolása a közzétett bejelentések és az engedélyezett szabadalmak adatait közlő rovatokban. Az olvasó aligha tudja elképzelni, hogy ezt a közlönyt IBM-kompozereken, fáradságos manuális munkával szedték, melynek során minden betűtípusváltás gömbfejcserevel is járt (egy-egy ügy adatainak leírása során 4–5 alkalommal kell fejet cserélni). Kompozereink 1990-re fizikailag és erkölcsileg teljesen elkoptak azáltal, hogy az IBM megszünteti az alkatrészellátást, és korábbi ügyfeleinek egységesen a DTP (desktop publishing) rendszerekre való átállást kínálja. Nincs ez másképp a mi esetünkben sem. A nyomda "elektronizálása" azt jelenti, hogy az 1990 januárjában megjelent közlöny "fotoeredetijét" már egy lézerprinterrel együttműködő szedő-tördelő program állította elő.

Alapos mérlegelés után döntöttünk a T<sub>E</sub>X-program mellett, amely ugyan – szemben a szóba jövő VENTURA vagy PAGEMAKER programokkal – nem képes WYSIWYG ("azt látod, amit majd kapsz") megjelenítésre, viszont hajlékony, a speciális műszaki igényeknek kiválóan megfelel, és képes tipográfiai előírásaink programozott kezelésére is. A T<sub>E</sub>X szedő-tördelő programot a 70-es évek közepén Donald Knuth fejlesztette ki, nevét a görög "művészet" szóból kapta. A megjelenítéshez a szövegközi szövegformázó parancsok és vezérlések egész arzenálját használja. A nagy szolgáltatóközpontok számos jól ismert nagy adatbázisa is T<sub>E</sub>X-alapú szedésből származik (ez például egy Dialog által offline nyomtatott lapokon szereplő furcsa jelek sorozatából az értök számára gyorsan kiderül).

Minthogy e parancsok a COLT adatbázist kezelő TEXTAR megjelenítő, formátumleíró nyelvnek segít-

ségével, illetve TEXTAR-alapú célprogramokkal könnyedén elhelyezhetők a nyomdába kerülő "forrás-szövegben", a T<sub>E</sub>X és a COLT együttműködésével a nyomdászoknak egy-egy rovat tördelésekor sokszor csak ellenőrző szerep jut.

A nyomda belépése az elektronikus adatellátásba komoly kérdéseket vetett fel, nevezetesen, hogy a szedés folyamatosan történjék-e, a megkorrektúrázott szöveget pedig – ismerte az állomány szerkezetét – fejtsük vissza adatbázisba (számos hivatal dolgozik így), vagy fordítva: adatbázis biztosítsa-e a nyomda adatellátását (a jól elektronizált hivatalok nagyobb része ezt a módszert követi). Mi végül is a második módszer mellett döntöttünk. A nyomda bibliográfiai adatokkal való ellátását egy e célra (is) szolgáló "hatlővetű" adatbázis, a COLT biztosítja, melyet TEXTAR alapon definiáltunk. A COLT, amely egyebek között az OLT és a SZKIV, valamint a LAJSTROM előretékinő adatellátását is megoldja (közlönyünk adatminőségének "jóságával"), az INPADOC-nak szóló magyar adatszolgáltatás feladatát látja el (a későbbiekben a szabadalmi leírásokat kísérő angol nyelvű kivonatok nívós megjelenítését, esetleg a leírások előlapjainak kinyomtatását ugyancsak támogatni fogja).

A Közlöny jósága itt nem akármit jelent. Az utólagos adatrögzítés minősége, nemkülönben saját ügyviteli rendszerünk adatminősége nem elégíti ki a közhasznú adatbázis építésének követelményeit. Az adatbázisok előállításának folyamatába – még ha ez nem is vallják be – a hazai adatbázis-előállítók zöme nem definiált szigorú adatellenőrzési modult (ezzel szemben az USA szabadalmi adatbázisának előállításában, az adatok átvételében a katonai szabványokban előírt mintavételezést és ellenőrzést vezették be). Számunkra a nyomdai korrektúra utáni állapot jelenti azt az adatminőséget, amelyet a továbbiakban elfogadhatónak tekintünk. Az adat-előállításban ígéretes a relációs szerkezetű TEXTAR-alapú adatbázis, amely lehetővé teszi, hogy egy-egy vállalat vagy egy-egy feltaláló neve (ha biztonsággal felismerhető, hogy egyazon vállalatról vagy személyről van szó) mindig egyazon formában kerüljön az adatbázisba. Bár az alkalmazott technika új, eddigi tapasztalataink több mint kedvezőek.

## Terveink – NSZO-5. védjegy, ipari minta adatbázis

Az előzőekben leírtak szerint megalapozott az az elképzelés, hogy saját tájékoztatói lehetőségeink fejlesztését a közhasznú tájékoztatás szempontjainak erős szem előtt tartásával érjük el. Az általunk elképzelt adatbázis ideáltípusa egy olyan bibliográfiai adatbázis, amely a közzétételt követő szakaszban szöveges kivonatokat is tartalmaz (a későbbiekben pedig az idegen nyelvű publikálás igényének is megfelel – legalább a címek angol és orosz nyelven történő megadásának erejéig). Ehhez az "electronic

capturing" szolgáltatást megoldást. "A kulcs tehát a nyomdában van." 1990 januárja óta új problémáink is keletkeztek; ezúttal nem azzal küszködünk, hogy hogyan szerezzünk adatrögzítő kapacitást, milyen formátumban adjunk át adatokat a számítástechnikai cég fejlesztőinek, hanem ténylegesen szakmai problémákra koncentrálnunk. A bibliográfiai adatbázis melléktermékként mindenestre már rendelkezésre áll.

Nem biztos azonban, hogy itt meg kellene állnunk. A magyar iparjogvédelem "pszichológiai határhoz" közeledik. Rövidesen elérjük a 200 000. lajstromszámot. Jelenleg a szabadalmi leírások kinyomtatását a SZIK nyomdai alvállalkozók bevonásával végzi (kevesen tudják, hogy a leírások kinyomtatása, ami végül is a hatósági tevékenység elengedhetetlen része, a költségvetés három legnagyobb költségelemének egyike). Eltökélt szándékunk, hogy a 200 000-nél magasabb lajstromszámú leírások kinyomtatására már csak olyan alvállalkozóknak adjunk megbízást, akik DTP-eszközökkel rendelkeznek, és a nyomtatott leírásokon kívül a szedésből származó ASCII-, számítógéppel olvasható szöveggállományok átadására is vállalkoznak.

Ez a megoldás a teljes szövegű adatbázisok irányába mutató kezdeményezés. Forrásokkal való ellátottságunk, de akár az adatok születésének tempója is egyelőre inkább csak a megfelelő tárolási és megjelenítési mód kialakítását kívánja meg, a későbbiekben azonban merészebb tervek is dőlgethetünk majd. A világ nagy, tökeerős szabadalmi hivatalai a faksimile képtárolással látták megoldottnak óriási papírchívumaik elektronikus leképezését. Ezek konkurens elérésű adattárak, szemben a szabadalmi táruk gyűjteményeivel. Az elektronikus, optikai lemezes archívumokban tárolt leírásokat egyszerre, egy időben többen is használhatják. Ne feledjük azonban, hogy ezek még nem igazi teljes szövegű adatbázisok, bennük az adatok hozzáférése a bibliográfiai adatbázisok adatmélységének felel meg.

Számunkra a CD-lemezes faksimile képtárolás még jó ideig álom marad, a szabadalmi leírások teljes szövegű adattárolása azonban az optikai lemezes

adattárolás más módszereinek igénybevételével reális lehet. Ehhez a ma már széles körben elterjedt, egyszer írható (WORM) lemezes optikai meghajtók használata látszik célszerűnek. A WORM-lemezes kísérletek az integrált kép- és szövegtárolás megvalósítása felé vezetnek. Bizonyos alkalmazásokban ugyanis a képtárolás már nem nélkülözhető. Ilyen lehet például a védjegyek vagy az ipari minták adattárolásának létrehozása.

Az OTH az I<sup>2</sup>F második ütemében három adatbázis építésére vállalkozik. Ezek közül kettő a nemzeti úton bejelentett védjegyek, illetve ipari minták bibliográfiai adatainak 10–10 éves archív állományait tartalmazza majd a COLT-hoz hasonló, TEXTAR alapú adatbázis-struktúrában. Itt az adatok frissítését is – ahogyan ez elvárható – az "electronic capturing", a Szabadalmi Közlöny és Védjegyértesítő kiadása fogja biztosítani.

Külön figyelmet érdemelnek az osztályozási, nevezéktani adatbázisok. Ezek a bibliográfiai adatbázisokban végzett keresések támogatásában, így a hatósági munkában is nélkülözhetetlenek. Közülük a legfontosabb az NSZO, melynek nyomdai előállításához a VENTURA programot vettük igénybe. A szigorúan strukturáltan szedett állomány adatbázisba fejtevése révén a nemzetközi adatbázispiacra elérhető INPI-4E adatbázishoz hasonló magyar nyelvű adatbázis állhat majd az I<sup>2</sup>F-hálózat használóinak rendelkezésére. Hasonló, de összehasonlíthatatlanul egyszerűbb feladatot jelent a Nizzai megállapodás szerinti mutató (ez a védjegyek nemzetközi osztályozásának mutatója), valamint a Locarnói megállapodás szerinti mutató (ez utóbbi az ipari minták osztályozásának nemzetközi mutatója) adatbázisba szervezése. Ezek létrehozása előrehaladott stádiumban van.

A sok elképzelésnek és tervnek végül is egy hatékony és integrált szabadalmi/iparjogvédelmi tájékoztató rendszerben kell megtestesülnie. Ehhez hitünk szerint megfelelő eszközeink, programjaink vannak. A többi pedig már rajtunk áll.

Beérkezett: 1990. I. 26-án.

## Egy önfenntartó információs szolgálat

Az NTIS (USA) naponta több mint 20 000 információ termékét ad postára, évente több mint 26 millió referátumot dolgoz fel, és minden évben több mint 60 000 új dokumentummal (kutatási jelentésekkel, szabadalmi alkalmazásokkal stb.) fejleszti állományát. Éves költségvetése 22 millió dollár, s teljes mértékben önfenntartó.

/Information Hotline, 21. köt. 2. sz. 1989. p. 1./

(Papp István)

## Kémiai szakirodalom

1988-ban Japán produkálta a kémiai publikációk 20%-át és a kémiai szabadalmak 54%-át. Az amerikai kiadvány- és szabadalomtermés negyedrészt, a szovjet anyag 12%-át teszi ki az egésznek. 1988-ban 46 nyelven (az angol foglalja el a vezető helyet) jelentek meg kivonatolt folyóiratcikkek a kémia területén. A biokémia jelentkezett a legtöbb publikációval: az egész termés 34,3%-ával.

/Information Hotline, 21. köt. 7. sz. 1989. p. 1./

(Papp István)