

## ADATBÁZIS-ÉPÍTÉS: PÉLDÁUL AZ INSPEC

Válas György

Országos Műszaki Információs Központ és Könyvtár

### Bevezetés

Magyarországon is égetően időszerűvé vált az online kereshető hazai adatbázisok létrehozása és ezzel együtt a meglévő saját referáló folyóirataink számítógéppel segített szerkesztése. Érdekes tehát körülnéznünk a nagyvilágban, hogy hogyan kell ezeket a feladatokat megoldani. Ha adottságaink, lehetőségeink sok mindenben nem is mérhetők a legjobbakéhoz, tanulunk akkor is tőlük célszerű, természetesen gondosan mérlegelve, hogy a tapasztalataikból mi az, amit változtatás nélkül átvehetünk, mi az, amit csak megfelelően adaptálva, és mi az, amit csak távlati fejlesztési célként vehetünk figyelembe.

Az INSPEC (**I**nformation **S**ervices for the **P**hysics and **E**ngineering **C**ommunities = információs szolgálat a fizikusi és mérnöki közösségek számára) a legjobbak között van, ezt világszerte elismerik. Ez vonatkozik mind az adatbázisra, mind pedig a neki megfelelő referáló folyóiratokra (*Science Abstracts*). Tőlük tehát van mit tanulnunk, különösen akkor, ha éppen bibliográfiai adatbázist akarunk létrehozni. Nézzük végig lépésről lépésre, hogyan készül az INSPEC! Vizsgálatunkban egyrészt az INSPEC-kel való sokéves ismeretségünk lesz segítségünkre, másrészt a személyes konzultációk, amelyekre egy 1985. őszi tanulmányút szolgáltatott alkalmat, harmadrészt az INSPEC rendelkezésünkre álló dokumentációs anyagai, amelyek egy részéhez szintén az említett tanulmányúton jutottunk. A leírásban olyan részletességre törekszünk, hogy aki bibliográfiai adatbázis építésébe fog, az lépésről lépésre összehasonlíthassa az elképzeléseit az INSPEC-ben kialakult gyakorlattal.

### Az adatbázis jellege, témája, mérete

Egy adatbázis tervezését annak meghatározásával kell kezdeni, hogy mi legyen benne. Ez persze a leg-szorosabban és kétoldalúan összefügg az adatbázis méretével. Ha elhatározunk egy összetételt, az meghatározza a méretet. Ha ez a méret meghaladja a lehetőségeinket, akkor vagy kompromisszumra kényszerülünk a tartalommal illetően, vagy meg kell találnunk az utat lehetőségeink bővítésére.

A tervezésnek ez az első lépése az INSPEC esetében nagyon egyszerű volt. Ez az adatbázis ugyanis meglévő, közel háromnegyed évszázada megjelenő referáló folyóiratok számítógépre viteléből született. Így eleve adva volt az alapjellege: bibliográfiai adatbázis, valamint a tematikája: a meglévő referáló lapok témaköre.

Ebből nagyon pontosan fel lehetett mérni a méretet is, hiszen nemcsak a havonta referált dokumentumok száma volt adott, de még a tartalmi kivonatok terjedelmének is megvolt a jól bejárodott hagyománya.

A Science Abstracts referáló folyóiratot 1898-ban indította el a brit elektromérnökök társadalmi szervezete, az *Institution of Electrical Engineers (IEE)*. Akkori témaköre az elektromosság és a fizika volt. A folyóirat azután hamarosan kettévált, megindult külön a *Physics Abstracts* és külön az *Electrical Abstracts*. Ez utóbbi később felvette az *Electrical and Electronics Abstracts* nevet. Ebből 1966-ban vált ki és lett önálló referáló folyóirattá a *Computer and Control Abstracts*. 1969-ben, nem sokkal a két úttörő, az *Index Medicus* és a *Chemical*

Abstracts referáló folyóiratok számítógépes változata, a MEDLARS, illetve a CA Search után, tehát a legkorábbi számítógépes bibliográfiai adatbázisok sorában született meg a Science Abstracts csoportba tartozó három referáló folyóirat anyagából a háromszekciós INSPEC adatbázis, akkor még csak mágnesszalagos formában. Az online szolgáltatóközpontok megszületésekor azután az online szolgáltatott adatbázisok sorában is a legelső és a legnépszerűbbek közé került. Jelenleg 7 nemzetközi és 3 országos nyilvános online szolgáltatóközpontban érhető el üzemszerűen az INSPEC adatbázis. (Nem tekintve üzemszerűnek az adatbázisokat órarend szerint váltogató, a hét nem mindegyik munkanapján működő és nem teljesen nyilvános moszkvai, szófiai és prágai szolgáltatást.) Ezenkívül számos helyi online központ is futtatja.

A Nagy-Britanniában az információs technika évének nyilvánított 1983-as évben indult meg – részben az Electrical and Electronics Abstracts, részben a Computer and Control Abstracts anyagából önállósulva – a folyóiratcsalád negyedik tagja, az IT Focus és vele az INSPEC adatbázis negyedik szekciója. Így alakultak ki a jelenlegi négysektori INSPEC mai témakörei:

- fizika,
- elektromosság és elektronika,
- számítástechnika és irányítás,
- irodai információs technika és telematika.

Bibliográfiai adatbázisnál a referált témakörök körülhatárolása után a következő lépés a figyelt dokumentumtípusok kijelölése és a figyelési kör szélességének a megtervezése. Az INSPEC esetében ezt is készen adta a referáló folyóiratok szerkesztési gyakorlata.

Az INSPEC által referált két fő dokumentumtípus a folyóiratcikk és a konferenciaanyag. Ebben a két dokumentumtípusban teljességre törekszenek. A teljesség persze relatív fogalom. Igyekeznek referálni minden olyan folyóiratot és konferenciát, amelynek fő vagy egyik fő témája az adatbázis valamely témaköre, ezenkívül minél több olyan folyóiratot, amelyben viszonylag gyakran találhatók a figyelt témakörökről készült cikkek. Ez utóbbiak között nagyon fontos multidiszciplináris folyóiratok is vannak, például a *Nature*. Végeredményben az INSPEC összes adatrekordjának 80%-át teszik ki a folyóiratcikkek, 15%-át a konferenciaanyagok.

A többi dokumentumtípusban nem céljuk a teljesség. Az évi 1000 könyvreferátum azokról a könyvekről készül, amelyeket a kiadójuk elküld az INSPEC-nek. Induláskor az INSPEC szabadalmakat is referált, de ezt a dokumentumtípust 1977-től kizárták az adatbázisból. A kutatási jelentések referálása legfőképpen kutatóintézetek jelentéssoro-

zataira szorítkozik (köztük van például a magyar KFKI – Központi Fizikai Kutató Intézet – jelentéssorozata); a disszertációk referálása teljesen esetleges.

A földrajzi és a nyelvek szerinti lefedésben az INSPEC az adott témakörök teljes világirodalmának a figyelésére törekszik. Így a referált dokumentumok 84%-a angol, 4,1%-a orosz, 4,1%-a német, 1,9%-a francia, 1,8%-a japán nyelvű; a többi nyelv, köztük a magyar, együttesen 3,9%-ot tesz ki.

A dokumentumokat kiadó országok szerinti megoszlás: a referált anyag 35,5%-a származik az USA-ból, 13%-a a Szovjetunióból, 8,5%-a Japánból, 8%-a együttesen az NSZK-ból és az NDK-ból, ugyancsak 8%-a Nagy-Britanniából, 5%-a Franciaországból és 22%-a a többi országból, beleértve a nagyon sok folyóiratot kiadó Hollandiát. Az INSPEC összesen körülbelül 60 ország kiadványait referálja. Mind a nyelvek szerinti, mind az országok szerinti megoszlás lényegében tükrözi a figyelt témák világirodalmának tényleges megoszlását. (A magyar folyóiratok közül a *Tudományos és Műszaki Tájékoztató* is referálja az INSPEC, így nagy az esély rá, hogy ez a dolgozat is bekerül a referált cikkek sorába.)

A felsorolt szerkesztési szempontok már meghatározzák az adatbázis méretét. A fizika szekció évi 130 000, az elektromos és elektronikai szekció évi 67 000, a számítástechnikai és irányítási szekció évi 54 000, az irodai információs technikai és telematikai szekció évi 3000 dokumentumot referál. Az átfedések miatt mindez együttesen valamivel több mint 200 000 adatrekordot jelent évente. 1969 óta az adatbázisban mintegy 2,5 millió adatrekord gyűlt össze.

## Szervezeti és számítástechnikai háttér

Vizsgáljuk meg, hogy a felvázolt adatbázis létrehozására milyen szervezet és milyen eszközök állnak rendelkezésre!

Az IEE 1871-ben alakult. Teljes egészében az egyéni tagdíjakból és kiadványai bevételeiből tartja fenn magát. Az állami támogatás abból áll, hogy nem profitorientált szervezet lévén, adót nem fizet. Ez alól eddig egy kivétel volt: 1969-ben jelentős egyszeri állami támogatást kapott számítógépesítésre. Ez a támogatás tette lehetővé a számítógépes INSPEC adatbázis létrehozását.

Az INSPEC az IEE önálló részlegeként 1967-ben alakult. Egyike annak a három szervezetnek, amely az IEE kiadványait állítja elő. (A másik kettő egy kiadóvállalat, amely főleg elsődleges tudományos-

műszaki folyóiratokat és konferenciakiadványokat ad ki, és a *Peter Peregrinus Ltd.*, amely nyomdát tart fenn, és bér munkában végez adatbázis-építést.) Az INSPEC-részleg feladatköre az ISPEC adatbázis és a Science Abstracts referáló folyóiratok előállítás, az EMIS adatbázis és a hozzá kapcsolódó nyomtatott kiadványok létrehozása, az INSPEC adatbázis anyagából válogatott *Current Papers* szignaletikus és a *Key Abstracts* referáló kiadványok kiadása, adatbázis-építés bér munkában, valamint szelektív információterjesztés az INSPEC adatbázisból egyedi és csoportos (standard) keresőprofilokkal. (Ezekről részletesen fogunk szólni.) Mindezt úgy kell végeznie, hogy az ezekből származó tiszta nyereségével jelentős mértékben hozzájáruljon az IEE fenntartásához.

A felsorolt munkákat körülbelül 140 fő gárdának kell elvégeznie. Ebbe nem számítanak bele azok a gazdasági és gondnoksági dolgozók, akik az IEE állományában végzik a munkájukat; velük együtt az összlétszám mintegy 250 fő.

Az INSPEC hozzávetőleg 140-es létszámával három részlegre tagozódik. A szerződéses munkák részlege (EMIS adatbázis és adatbázis-építés bér munkában) és a rendszerfejlesztési részleg (amelyhez a programozók is tartoznak) nagyon kicsi. A létszám zöme, nagyjából 120-an az igazgatóhelyettes közvetlen vezetése alatt álló és négy osztályra tagozódó adatbázis-előállító részlegbe tartoznak. Ez a részleg állítja elő az INSPEC adatbázist és a hozzá tartozó nyomtatott kiadványokat (Science Abstracts, Current Papers és Key Abstracts), valamint ez végzi a szelektív információterjesztést. (Történetileg ugyan a Science Abstracts volt az elsődleges és az abból kialakított INSPEC a másodlagos, de ma már inkább a referáló folyóirat tekinthető az adatbázis melléktermékének.)

Az adatbázis-előállító részleg legnagyobb osztálya 45–50 fővel a szerkesztő-indexelő osztály. Feladata a referáló dokumentumok kiválasztása, kivonatolása és indexelése, de ők gondozzák a tezauszst, az osztályozási rendszert és a figyelt folyóiratok jegyzékét is.

A termelési osztály végzi a referálandó dokumentumok beszerzését, a korrektúrát, a segéd-adatállományok (folyóirat-adatállomány, konferencia-adatállomány, névadatállomány stb.) gondozását, a deskriptív katalogizálást és a hozzá csatlakozó első adatrögzítést. Ők tartják a kapcsolatot a külső vállalatokkal (adatrögzítés és nyomda). Hozájuk tartozik a gyorsmásolás is.

A marketingosztály feladata minden, ami a felhasználókkal való kapcsolathoz tartozik, tehát a szoros értelemben vett marketingmunkán kívül a

szelektív információterjesztés, az adatbázis félhavi növekményét tartalmazó mágnesszalagok előállítása és eljuttatása az előfizetőkhez, a nyomtatott termékek terjesztése, az INSPEC-et "pörgető" online szolgáltatókkal való kapcsolattartás és a felhasználók oktatása is.

A rendszerüzemeltető osztály a számítógéprendszert üzemelteti. Ide tartoznak az operátorok, a hardveres szakemberek és az üzemeltető szoftveres munkatársak.

Az INSPEC előállításának hardverbázisa kezdetben az (akkor nagy közepesnek számító) ICL 2950/10 típusú számítógép volt, amelyet 1969-ben állami támogatásból vásároltak. Ehhez akkor 0,5 Mbájt központi tár és 4 db 7 sávú mágnesszalag-egység tartozott, valamint egy 1500 sor/perc sebességű sornyomatató és gyors papírszalag-olvasó. Ezt azóta 3x200 Mbájt cserélhető mágneslemezes háttértárral látták el, és 3 db 9 sávú mágnesszalag-egységgel egészítették ki. Ez az ICL-gép azóta is működik, de a feldolgozást fokozatosan átvizik az új, nemrég üzembe állított rendszerre. Az öreg gépmatuzsálem feladata marad az IEE adminisztratív-gazdasági munkáinak kötegelte feldolgozása és a mágnesszalag-előállítás néhány alrendszere (a 7 sávú és a 800 bpi írássűrűségű szalagok előállítása). Átmenetileg az ICL-en fut a fényszedést előkészítő program is.

A jelenlegi fő számítógérendszer két VAX 11/750 típusú, megamini kategóriájú számítógépből, néhány személyi számítógépből és terminálból áll, DECNET-hálózatban egymással összekapcsolva. Ezenkívül néhány, a hálózatba be nem kapcsolt személyi számítógép szolgál célfeladatokra (pl. az online információkeresésre és a mágnesszalagos szolgáltatás adminisztrációjára).

A két VAX 11/750 közül az egyiknek 2 Mbájt központi tára, 80 Mbájt fix és 2x67 Mbájt cserélhető "saját" mágneslemezes tára van. A nappali interaktív üzemmódban csak ez a VAX működik, a másik csak az éjszakai kötegelte feldolgozásba kapcsolódik be. Ennek 3 1/4 Mbájt központi tára, 80 Mbájt fix és 3x67 Mbájt cserélhető "saját" mágneslemezes háttértára van. A "saját" háttértáron kívül még 5x256 Mbájt cserélhető mágneslemezt együttesen használhat a két gép. Ugyancsak a két VAX együttes használatára szolgál 6 db 9 sávú mágnesszalag-egység, egy 1000 sor/perc sebességű sornyomatató, egy mátrixnyomatató, egy 600 sor/perc sebességű printer-plotter, egy VAX 100 grafikus terminál nagy felbontású képernyővel és "egér"-rel, 2 modem a távolsági összeköttetésekhez, 30 alfanumerikus terminál és egy Alphaword optikai karakterolvasó (OCR) készülék. A berendezések egy részének (OCR, modemek) nincs szerepe az

INSPEC adatbázis és a nyomtatott kiadványok jelenlegi előállítási rendszerében.

Az INSPEC saját belső erőforrásai mellett két területen vesz igénybe jelentős külső erőforrást: az adatrögzítés túlnyomó részét külső szolgáltatóvállalatokkal végeztetik (három ilyen vállalattal van szerződésük), a fényszedést és a nyomtatást-kötést pedig szakvállalattal.

## Rekordszerkezet

Az adatbázis tervezésének kezdeti lépései közé tartozik a rekordszerkezet meghatározása. Ez két részből áll. Egyrészt azt kell eldönteni, hogy milyen adatmezőket tartalmazzon a rekord, másrészt azt, hogy ezek a mezők hogy helyezkednek majd el a rekordban, és milyen jelöléssel, vagyis a számítástechnikai értelemben vett rekordformátumot.

Tulajdonképpen nem egy, hanem három rekordszerkezetet és ezen belül rekordformátumot kell megtervezni: egy belsőt, amelyet a szerkesztést támogató rendszer használ, egy külsőt, amely a kiküldendő mágnesszalagokra érvényes, végül egyet a nyomtatott változat fényszedése számára.

A belső rekord a két külső rekord valamennyi adatmezője mellett még a szerkesztést segítő rendszer működéséhez szükséges segédmezőket is tartalmazza. Az INSPEC-nél ezek jelzőkarakterek, amelyek egyrészt azt tartják nyilván, hogy az adott rekord a szerkesztés melyik szakaszánál tart, másrészt azt, hogy a kész rekord a végtermékek közül melyekbe került már bele. A belső rekordformátum teljesen a feldolgozó rendszer belügye, a használt adatbázis-kezelő rendszer (itt az RMS) szabja meg.

A fényszedés (meglehetősen bonyolult) rekordszerkezetét az INSPEC esetében egyértelműen megszabta a nyomtatott kiadványok hagyományos formátuma, amelyen nem kívántak változtatni. A rekordformátumot a fényszedő szakvállalat írja elő.

Külső rekordformátum gyanánt általában a szabványos vagy egyezményes formátumok közül szokás választani. Az INSPEC-re 1973 után az ISO 2709 formátumot (MARC-formátum) választották. Ebben a formátumban minden adatmező változó hosszúságú, és több adatelemből állhat. (Adatelemből például a szerzők adatmezőben egy szerző neve.) Az adatelemből egy adatelemből határoló karakterből (delimiter) és magából az adatból áll. Az adatmező egy jelzőkarakterrel (változatszám) kezdődik, ezután következnek sorban az adatelemből, végül egy adatmező-határoló karakter. A rekordfej után elhelyezett és fix hosszúságú elemekből álló mutató (directory) őrzi minden adatmező hívójelét (tag), hosszát és kezdőpozícióját.

A rekordformátumot teljes részletességgel és pontossággal írja le a *Tape Services Manual* című kiadvány. Ez repülőlapos kivitelű, hogy az esetleges változásokat azonnal át lehessen vezetni benne.

A rekordformátumnál is gondosabb megfontolást igényel, hogy mi kerüljön bele a rekordba. Ez ugyanis alapvetően meghatározza az adatbázis használhatóságát. Ami kimarad, vagy rossz koncepcióval kerül bele, a felhasználónál már nem hozható helyre. Ami feleslegesen kerül bele, az legalábbis indokolatlanul növeli a költségeket.

Az INSPEC esetében az adatmezők tartalmának egy része már többé-kevésbé adott volt a nyomtatott változat hagyományával. A bibliográfiai adatokról például csak azt kellett eldönteni, hogy melyik legyen belőlük önálló mező, melyik nem, figyelembe véve, hogy ami külön mezőben van, az a további feldolgozás során még összevonható, ami viszont egyben van, az már nem választható szét. Ezek a hagyományos adatmezők azonban még nem alkotnak gépi keresésre alkalmas adatrekordot. A *SCISEARCH* adatbázisban például, amely csak bibliográfiai adatmezőket tartalmaz, alig lehet téma szerint keresni. Ki kellett egészíteni az INSPEC adatrekordot a gépi keresést lehetővé tevő mezőkkel (deskriptorok, tárgyszavak), a keresési lehetőségeket bővítő mezőkkel (nyelv, dokumentumtípus stb.), továbbá be kellett vinni minden rekordba a nyomtatott változatban csak fejezetcímként szereplő osztályozási jelzetet.

Vegyük hát sorra az INSPEC adatbázis legfontosabb adatmezőit! Éppen ezeknek az adatmezőknek a sikeres megtervezése az egyik legfontosabb tényezője annak, hogy az INSPEC a legnépszerűbb adatbázisok közé tartozik a világon.

A hagyományos bibliográfiai (és korábban a szabadalmi) adatok a legnagyobb részletességgel külön-külön adatmezőbe kerülnek. Így az INSPEC-en alapuló információs szolgáltatásban a szolgáltató döntheti el, hogy ezek közül melyeket tesz kereshetővé. Egyetlen kivétel van: a folyóirat kötete és száma egy adatmezőbe kerül.

A hagyományos bibliográfiai adatok mellett az adatbázisba bevettek olyan nem hagyományos bibliográfiai adatokat is, amelyek a nyomtatott változatba nem fértek bele (CODEN, ISSN, ISBN stb.). Külön adatmezőben módot hagytak a nem hagyományos hordozó (pl. mágnesszalag vagy mikrofilmlap) leírására is.

A bibliográfiai adatok között a folyóiratcím mindig szabványos módon rövidített formában szerepel. A rövidítésből, a CODEN-ből vagy az ISSN-ből a teljes cím, illetve a teljes címből a rövidítés, a CODEN és az ISSN kikereshetők az *INSPEC List of Journals and Other Serial Sources* című kiadványból.

Az angol nyelvű dokumentumok címét eredeti formájukban, a nem angol nyelvűeket angol fordításban viszik be az adatbázisba. A nagyon semmitmondó címeket kivételesen ki is egészíthetik, szögletes zárójelek közé téve a kiegészítést.

Valamennyi, az eredeti dokumentumon feltüntetett szerző neve bekerül az adatbázisba, akkor is, ha akár száz szerzője is van a dokumentumnak, ami a kísérleti részecskefizikában nem számít meglepőnek. A keresztnéveket viszont mindig kezdőbetűkkel helyettesítik. Ugyanígy járnak el a szerkesztőkkel és a fordítókkal is.

Csak az első szerző munkahelyét tüntetik fel, ha az a primer dokumentumon megtalálható.

Az adatbázis rekordjai és a Science Abstracts referátumai (egyes tételei, amelyek a bibliográfiai adatokat és a kivonatokat tartalmazzák) között egyértelmű kapcsolatot teremt a referátumsorszám. Ha egy dokumentum témája olyan, hogy több referáló folyóiratba is bekerül a Science Abstracts négy sorozata közül, akkor a megfelelő adatrekord több referátumsorszámot is tartalmaz. A referátumsorszám tartalmazza a megfelelő sorozat szekciójelét és a referálás évének két utolsó számjegyét.

A gépi információkeresés legfontosabb eszköze az INSPEC-ben a *deszkriptormező*. Ez teljesen új eszköz a nyomtatott referáló folyóiratokhoz képest, ahol ilyesmire nem volt szükség, hiszen a nyomtatott anyagban való "gyalogos" keresés során a címből és a kivonattól értékelhető a dokumentum relevanciája. A gépi keresést viszont nem lehet kerek mondatokból vagy éppen mondatok sorából végezni (legalábbis a mesterséges intelligencia kutatásának mai szintjén nem). Itt tömör, a dokumentum tartalmát jól jellemző kifejezésekből kell kiindulni, ehhez pedig a szinonimák és kváziszinonimák lehetséges igen magas száma miatt gondoskodni kell az indexelő és a kereső gondolkodásának az összehangolásáról. Szükség van tehát a kötött szó- és kifejezésjegyzékre, de a leghasznosabb a kötött szójegyzék legszervezettebb változata, a *tezaurusz* (lásd később). A deszkriptorok nélküli adatbázisok (pl. SCISERACH, INPADOC, VINITI) mindig sokkal bizonytalanabb keresési eredményeket nyújtanak, mint a deszkriptorosok. Azokban az adatbázisokban sem kielégítőek a keresési eredmények, amelyek használnak ugyan deszkriptorokat, de adatrekordonként csak egy deszkriptort engednek meg (pl. *EI/Engineering Meetings*). Az INSPEC adatbázis-hoz nagyon jól szervezett, részletes tezauruszt készítettek, és azt rendszeresen, kétévenként fejlesztik. Az ebből a tezauruszból vett deszkriptorok korlátlan számban rendelhetők egy adatrekordhoz. Háromnál kevesebb deszkriptort ritkán találunk egy adatrekordon, gyakori viszont a tíz-tizenöt deszkrip-

torral indexelt dokumentum. Ez kiemelkedő helyet biztosít az INSPEC-nek a nagy adatbázisok között a keresés teljessége és pontossága terén.

A deszkriptorok mellett a téma szerinti gépi keresés másik fontos eszköze az INSPEC-ben a *szabad tárgyszavakat tartalmazó adatmező*. Ebben a szerző saját terminológiájából vett kifejezések és más, a tezauruszban nem található, de a dokumentum témáját jól jellemző kifejezések találhatók. Már említettük, hogy a szabályozatlan kifejezések nem biztosítanak kellő összhangot az indexelő és a kereső között, így önmagukban nem szolgálják kielégítően a keresést. Miért kell akkor a sokkal jobb keresési lehetőségeket nyújtó deszkriptorok mellett mégis szabad tárgyszavakat alkalmazni? Ennek négy oka van. Először: a kereső sokszor olyan szűken akarja körülhatárolni a témáját, hogy ez deszkriptorokkal nem lehetséges. A tezaurusz terjedelmét ugyanis nem lehet korlátlanul növelni, mert az már bizonytalanságot okozna az indexelésben (amint az például a nagyon részletes tezauruszú INIS adatbázisban gyakori), ezért nem lehet minden részletkérdésre deszkriptorokat kijelölni. Másodszor: a tezaurusz csak bizonyos időközönként, az INSPEC esetében kétévenként aktualizálható, a tudomány pedig folyamatosan és gyorsan fejlődik. Csak szabad tárgyszavakkal írhatók le azok az új fogalmak, amelyekre még nem született deszkriptor. Harmadszor: bizonyos kifejezéstípusok nem vehetők be a deszkriptorok közé, de a dokumentumok tárgyának pontos leírására szükségesek. Ilyen például annak a számítógépnek a pontos típusjele, amelynek a szoftverjével vagy alkalmazásával a cikk foglalkozik, vagy annak a szerves vegyületnek a neve, szervetlen vegyületnek a képlete, amely a cikkben szerepel. Negyedszer: a tezaurusz fejlesztésének a legjobb kiindulópontja, hogy az elmúlt időszakban milyen új kifejezések bizonyultak gyakornak vagy más szempontból fontosnak a szabad tárgyszavak között.

Az INSPEC szabad tárgyszavai között időnként feltűnnek a deszkriptorok megisméltései is. Ennek az az oka, hogy az INSPEC saját szelektív információterjesztésében nem lehet deszkriptorok szerint keresni.

A deszkriptorokhoz hasonlóan a szabad tárgyszavak száma sincs korlátozva, és ugyanúgy általában három fölött, nagyon gyakran tíz fölött van, mint a deszkriptoroké.

A téma szerinti keresés harmadik lényeges eszközéül az *osztályozási jelzetek* szolgálnak. Ha pontosan körül akarunk határolni keresésünkben egy szűk témakört, akkor arra a deszkriptorok és a szabad tárgyszavak adnak jó eszközt. Ha azonban nagy témakört akarunk kiválasztani, akkor ezekkel

nem megy, mert a hierarchikus deszkriptorcsaládokból mindig a legszűkebben releváns deszkriptorral indexelik a dokumentumot, és a szabad tárgyszavakat is úgy adják, hogy minél szűkebben legyen rá jellemző. Márpedig gyakori, hogy a keresés több szempontja között ilyen tág körök is vannak. Ilyenkor segít a hierarchikus osztályozási rendszer, amelynek bármelyik hierarchiaszintje szerint lehet keresni. Így elkerüljük, hogy a tág témakört jellemző valamennyi deszkriptort össze kelljen gyűjtenünk. Ha például azokra a dokumentumokra vagyunk kíváncsiak, amelyek a részecskefizikában alkalmazott képfeldolgozási módszerekkel foglalkoznak, akkor a képfeldolgozás deszkriptorával és a részecskefizika osztályozási jelzetével kereshetjük őket.

A nyomtatott referáló folyóiratokban az osztályozási rendszer csak arra szolgált, hogy a referátumokat a szerint csoportosítsák fejezetekbe és alfejezetekbe. Ehhez egyetlen osztályozási jelzetet kellett adni minden dokumentumnak. Esetlegesen egy-két további jelzet szolgálhatott a keresztutalások alapjául, de ezzel csínján kellett bánni, mert nagyon növelte a terjedelmet. A megváltozott cél folytán az adatbázisban általában több osztályozási jelzettel indexelnek, hogy a dokumentum a rá jellemző valamennyi lényeges szempont szerint kereshető legyen. Az adható osztályozási jelzetek számára az INSPEC-ben nem írtak elő felső korlátot. Az osztályozási jelzetek száma nem lehet kevesebb, mint ahány szekcióba a dokumentumot besorolják, de gyakran eléri a tízet is. Az osztályozás mindig a legalacsonyabb releváns szinten történik. Mivel keresni valamennyi hierarchiaszinten lehet, a legalacsonyabb szintű osztályozás adja a legsokoldalúbb lehetőséget a keresésre.

Az eddig felsorolt adatmezők egyrészt téma szerint teszik lehetővé a keresést, másrészt egyes bibliográfiai adatok (szerzők, folyóirat, a kiadás éve stb.) szerint. Az INSPEC további adatmezői bizonyos *kiegészítő* keresési lehetőségeket is nyújtanak. Ezek legtöbbször szinte minden bibliográfiai adatbázisban megtalálható. Ilyenek a dokumentumtípus, a dokumentum nyelve, a referálás (az adatbázisba való felvétel) időpontja. Van azonban egy olyan mező, amely az INSPEC "találmánya", de az utóbbi években egyre több adatbázis átveszi. Ez pedig a referált dokumentum *tárgyalásmódja*: kísérleti, elméleti, gyakorlati, bibliográfia, áttekintés, új termék leírása stb. Ez a mező olyan szempont szerint teszi lehetővé a keresést, amelyet a felhasználók gyakran igényelnek, de a hagyományos adatmezők segítségével nem kereshető. Ebben a mezőben is több jelzetet kaphat egy dokumentum, ha több kategóriába is beleillik.

A bibliográfiai adatbázisoknak fontos jellemzője, hogy tartalmazzák-e a referált dokumentum kivonatát. Az INSPEC kezdettől fogva tartalmaz kivonatot, mégpedig gyakorlatilag minden referált dokumentumról, ez alól nagyon ritka a kivétel.

## Osztályozási rendszer és tezaurusz

Egy bibliográfiai adatbázis tervezése során az első lépések között kell eldönteni, hogy az adatbázis milyen témaköröket fedjen le. Ebben a tervezési szakaszban azonban ezt a feladatot csak elnagyoltan lehet elvégezni. A tervezés egy későbbi szakaszában erre vissza kell térni, és az adatbázis témáit teljes pontossággal körül kell határolni, hogy minél kevesebb olyan határterület legyen, amelyet öltetszerűen hol figyelnek, hol nem. Ez a precíz körülhatárolás persze nem lehet sztatikus. A tudomány fejlődését folyamatosan követni kell az adatbázis tematikai határaival is. A témakörök precíz körülhatárolásának és a körülhatárolás folyamatos fejlesztésének a legjobb eszköze az adatbázis *osztályozási rendszere*. Valószínűleg ezt kell az osztályozási rendszer legfontosabb céljának tekintenünk, amellyel, hogy az osztályozás az adatbázisban kereső eszköz, a nyomtatott referáló folyóiratban pedig a rendszerezés, szerkesztés eszköze is. Az osztályozási rendszer valamennyi feladatköre hierarchikus, több szintű rendszerrel látható el a legelőnyösebben.

Az INSPEC osztályozási rendszere ötszintű hierarchiával rendelkezik. A legfelső szint megfelel az adatbázis négy szekciójának, azaz a Science Abstracts négy sorozatának. A következő szint olyan nagy területeket jelöl ki, mint például a részecskefizika, a számítógépes szoftver, a távközlés. A legalsó szinten az INSPEC osztályozási rendszere annyira részletes, hogy némelyik jelzete már-már egyenértékű valamelyik deszkriptorral.

Az osztályozási jelzetek tartalmi leírásának fontos részei a "lásd még" típusú keresztutalások és a témakör egyes részleteire vonatkozó kiegészítő megjegyzések. Ez utóbbiak vagy azt írják le, hogy a témakörbe még mely, nem triviálisan beletartozó területeket kell beleérteni, vagy azt, hogy mely részterületeket zárnak ki belőle. Ilyen kiegészítő megjegyzések és keresztutalások az alsó három osztályozási szinten vannak.

Az osztályozási jelzeteket a szerkesztés folyamán használt belső változatban ellenőrző karakterrel egészítik ki. Ez jó esélyt ad az adatrögzítési hibák és az indexelés során elkövetett elírások gépi kiszűrésére. Ezenkívül ebben a belső változatban a középső három szinten külön karakter jelzi, hogy az osztálynak vannak-e aláosztásai.

Az INSPEC osztályozási rendszere az adatbázis több mint másfél évtizedes története során kétszer módosult és jelentős területekkel bővült. Ezért a régebbi időszakokra is vonatkozó retrospektív keresést külön kötetben kiadott megfeleltetési táblázat segíti. Így nem kell mindegyik időszak osztályozási rendszerét párhuzamosan használni.

Mint már említettük, a jól szerkesztett, sok, de nem túl sok deskriptort tartalmazó *tezaurusz* nagyon lényeges eleme annak, hogy az INSPEC-et a világ legjobb adatbázisai sorában tartják számon. Bibliográfiai adatbázis esetén a tezaurusz minősége (egyáltalán a léte) nagyon fontos eleme az adatbázis minőségének.

Az INSPEC tezaurusza mintegy 5500 deskriptort és mintegy 4500 kizárt kifejezést (nem deskriptort) tartalmaz. Multihierarchikus szerkezetű, vagyis egy deskriptornak lehet több közvetlenül fölérendelt deskriptora, és még a legfelső szinten is több hierarchiába tartozhat. A hierarchikus csoportokat a tezaurusz egy külön része teljes egészükben közli, a szintek pontos jelölésével (hierarchikus rész). A tezaurusz főrésze (alfabetikus része) minden deskriptor szóbokránban közli a közvetlenül alárendelt és a közvetlenül fölérendelt deskriptorokat, valamint a legmagasabb fölérendelt vagy csúcshierarchikus deskriptort (deskriptorokat). Ez utóbbi ad módot a nyomtatott változatban a hierarchikus rész fellapozására, így az összes alárendelt és fölérendelt deskriptor gyors áttekintésére. A hierarchikus rész használata nélkül csak több lépésben volna mód az összes alá- és fölérendelt deskriptor összegyűjtésére.

Az alá- és fölérendelt deskriptorokon kívül a szóbokránban "lásd még" típusú keresztutalásokat találunk más deskriptorokra (rokon deskriptorok), valamint keresztutalásokat a nem deskriptorokra, az osztályozási rendszer megfelelő jelzeteire és a korábbi tezauruszváltozatokban használt deskriptorokra. Emellett sok deskriptorhoz a használatra vonatkozó szöveges magyarázó jegyzetet vagy szöveges keresztutalást is találunk. Mindez kiegészül a deskriptor bevezetésének dátumával, ami a retrospektív keresések szempontjából lényeges.

Az osztályozási rendszert és a tezauruszt mágnesszalagon is megkapják az adatbázis mágnesszalagos változatának előfizetői. Az online rendszerek üzemeltetői tehát online tezauruszt is létesíthetnek, ha erre szoftverrel felkészültek. Ugyanezen a mágnesszalagon még az adatbázisban figyelt időszaki kiadványok listája is megtalálható.

## Előválogatás, a feldolgozandó dokumentumok beszerzése

Amivel eddig foglalkoztunk, az még nem maga az adatbázis-építés volt, hanem csak az előkészítése. Az előkészületi szakaszból a tényleges adatbázis-építésbe vezet át az *előválogatás* és a *beszerzés*. Előválogatáson azoknak a folyóiratoknak, más periodikáknak, nem periodikus sorozatoknak, konferenciakiadványoknak, cikkgyűjteményeknek stb. a kiválasztását értjük, amelyekből majd a ténylegesen referálandó cikkeket, előadásokat, egyéb dokumentumokat kiválogatják. Ez bizonyos mértékben még az előkészületi szakaszhoz tartozik, hiszen a kiadványok beszerzése nagyrészt előfizetéssel vagy szerződéses cserével történik; bizonyos mértékben már az adatbázis-építéshez, hiszen a dokumentumok egy részének a beszerzése folyamatos kiválasztással és folyamatos munkamenetet igénylő egyedi vásárlásokkal vagy ajándékozásokkal, alkalmi cserékkel megy. Az INSPEC esetében az előválogatáshoz és beszerzéshez is biztos kiindulópontot jelentett, hogy az adatbázis működő referáló folyóiratokból nőtt ki.

Az előválogatás munkamenete szempontjából a figyelt dokumentumokat az INSPEC három csoportba osztja:

- ◆ folyóiratok, egyéb időszaki kiadványok,
- ◆ konferenciaanyagok,
- ◆ könyvek és egyéb dokumentumok.

A legegyszerűbb a munka a harmadik csoporttal. A könyvek és egyéb dokumentumok közül ugyanis az INSPEC csak azokból választja ki a referálandókat, amelyeket a kiadójuk, a szerzőjük vagy valaki más referálásra megküld nekik. Itt tehát az előválogatás lépése kimarad. Ezt a politikát persze az alapozza meg, hogy az INSPEC jól bevezetett, tekintélyes adatbázis. Így azután a könyvek kiadóinak sokszorosan megtérülő reklámkiadás, ha küldenek egy-egy példányt az INSPEC témáiba vágó könyveikből bedolgozásra. Jelentősen emelheti az eladható példányszámot, ha az INSPEC az adott könyvet elfogadja és referálja. A nem eladásra szánt dokumentumok (például a disszertációk vagy az évkönyvek) esetén viszont a kibocsátó intézmény presztízsét emeli, ha az INSPEC a munkát referálja.

A folyóiratok, periodikus kiadványok, sorozatok előválogatásának eredménye az állandó és folyamatosan karbantartott folyóiratlista, amely a szerkesztést segítő számítógépes rendszerben házi adatbázis formájában is megtalálható (lásd később), és amelyet a szerkesztő-indexelő osztály gondoz. Ez a lista jelenleg mintegy 3700 címet tartalmaz. (Köztük 30 a magyar folyóirat és sorozat; több, mint például az osztrák, a dán, a finn, a svéd vagy az izraeli.)

A folyóiratok között kiemelt helyet foglal el 736 (jelenleg). Ezek a "borítótól borítóig" referált folyóiratok. Hogy egy folyóirat ebbe a kategóriába kerüljön, kettős követelményt kell kielégítenie. Az egyik követelmény a nagyon magas színvonal. A cikkek szigorú szerkesztőségi szelekciója biztosíték arra, hogy csak referálásra érdemes írások jelennek meg a kiadványban. Ez persze nem jelent csupa korszakalkotó újdonságot, ilyen igénye egy referáló lapnak vagy adatbázisnak sem lehet. Olyan cikkeket kell válogatni, amelyek a szakemberek bizonyos körét feltétlenül érdeklik. A borítótól borítóig referált folyóiratok másik kiválasztási szempontja, hogy a témakörük teljes egészében az INSPEC témakörébe essen. Ennek a kiválasztási szempontnak például nem felel meg a Nature, ezért az INSPEC-ben nem referálható borítótól borítóig, bár az egyik legrangosabb, legszigorúbb szerkesztőségi szelekciót érvényesítő folyóirat. (A borítótól borítóig referált magyar folyóiratok: *Acta Physica Hungarica*, *Budavox Telecommunication Review*, *Elektrotechnika*, *Information Bulletin on Variable Stars* és *Problems of Control and Information Theory*.)

A listán szereplő folyóiratok 45%-át szerzik be előfizetéssel, a többit ajándékozás vagy csere révén.

A lista karbantartása kétoldalú. Az újonnan megjelenő folyóiratokról általában a kiadók értesítik az INSPEC-et, így módjuk van már az első számmal megkezdeni a referálást. Amelyik folyóiratban viszont hosszabb ideig nem találtak referálásra kiválasztható cikket, azt törlik a listáról. Törlik a megszűnő folyóiratokat is.

Volt idő, amikor a feldolgozás gyorsítása végett néhány folyóiratnál megpróbálkoztak azzal, hogy a cikkek referálását korrektúrapéldányból (kefelevonatból) végezték még a megjelenés előtt. Erről azonban le kellett mondaniuk. Gyakoriak voltak ugyanis a bennmaradt hibák és az utólagos változtatások, így a közlés pontatlanná vált. Különösen az oldalszámok változtak gyakran a kefelevonat készítésétől a megjelenésig. Most már visszatértek arra, hogy minden folyóiratot a megjelent példányból referáljanak.

### Belső nyilvántartó adatbázisok

A házi folyóirat-adatbázisnak, amelyből általában kétévenként készítenek nyomtatott kiadványt *INSPEC List of Journals and Other Serial Sources* címen, hármas célja van.

Az első, hogy könnyűvé, egyszerűvé tegye a lista karbantartását. Az adatbázisban nagyon egyszerű az új adatrekordok felvétele, a fölöslegessé vált rekordok törlése, az esetleges cím- vagy egyéb változások

átvezetése, az esetleges fölfedezett hibák javítása. Ez a karbantartás online terminálról történik. Az adatbázisból teljesen automatizáltan, a szerkesztést segítő programrendszer programjaival készítik a nyomtatott folyóiratlista fényszedő üzembe küldhető mágnesszalagját. Az adatrekordok megfelelő adatmezőit külön erre a célra írt programmal válogatják ki, a kiadvány egyes részeihez külön-külön. Ezután rendezéssel készítik el ezeket a részeket: a teljes folyóiratcím szerint rendezett főrészt, valamint a rövidített folyóiratcím szerinti, a CODEN szerinti, az ISSN szerinti és a kiadó országok szerinti mutatókat. A nyomdai előkészítésre a szerkesztést segítő programrendszernek az az alrendszer szolgál, amely rendkívüli sokoldalúsága, rugalmas paraméterezhetősége révén az INSPEC rendszeresen megjelenő kiadványai mellett egyedi kiadványok szerkesztésére is alkalmas. A programok paraméterezése online történik. Ezzel az alrendszerrel oldalakra és hasábokra osztják a szöveget, elkészítik a fejezetcímeket, az egyedi és az állandó fejléceket, a lábjegyzeteket, oldalszámozást generálnak, kialakítják a nyomtatott rekord- (bekezdés-) formátumot, beleértve a magyarító adatmezőjelzések beiktatását, végül kijelölik a betűtípusokat és betűméreteket. A végeredmény olyan mágnesszalag, amely már közvetlenül betehető a fényszedő gépbe.

A belső folyóirat-adatbázis másik célja, hogy – ugyancsak online, menürendszerű hozzáféréssel – a folyóiratszámoké nyilvántartást vezesse. Ebben a folyóiratszámok sorsát az előfizetéstől az esetleges reklamációkon, a beérkezésem, a referálandó cikkek kiválogatásán, deskriptív katalogizálásán és az ehhez kapcsolódó első adatrögzítésen át az indexelési kötegekre való szétszedésig tartják nyilván. (Erről később még lesz szó.) A szétszedés után a folyóiratszámoké nyilvántartás helyét a dokumentumonkénti nyilvántartás foglalja el, az pedig már másik adatbázisban történik.

A folyóiratszámoké nyilvántartást végző alrendszernek egy hetenként egyszer futó programja automatikusan generál reklamáló levelet akkor, ha egy folyóiratszám úgy érkezett be, hogy az előző számok közül egy vagy több hiányzik, illetve ha a folyóirat kiadási gyakoriságához képest feltűnően hosszú ideje nem érkezett be a következő szám. Ezeket a reklamáló leveleket az alrendszer nyilvántartja, hogy ugyanarról a hiányról ne menjen ki közvetlenül egymás után több reklamáció. Ugyanez az alrendszer a beérkező folyóiratszámok közül kiszűri a duplumokat, nehogy egy cikket többször dolgozzanak fel.

A belső folyóirat-adatbázis harmadik célja, hogy tárolja azokat az adatokat, amelyek a folyóiratról az INSPEC adatbázis adatrekordjaiba bekerülnek.



Ezek az adatok a rekordokba nem közvetlen adatrögzítéssel jutnak be. A deskriptív katalogizáláshoz kapcsolódó első (interaktív) adatrögzítéskor a folyóirat-adatbázis megfelelő rekordjából hívják le az ilyen adatokat. Ehhez elegendő a folyóirat azonosítására alkalmas adatok egyikét begépelni, általában az ISSN-t. A képernyőn ellenőrzésképpen megjelennek a folyóirat adatai, majd ha az adatrögzítő jóváhagyja az azonosságot, akkor az összes szükséges adat automatikusan bekerül a dokumentum adatrögzítésébe. Ezzel nemcsak rengeteg adatrögzítési munkát lehet megtakarítani, hanem a hibalehetőségek száma is jelentősen csökken.

A folyóirat adatainak a begépelése rendszerint nagyon leegyszerűsített formában végezhető. Elég például azt beírni, hogy VOL7NO11, ebből a program automatikusan kialakítja a helyes alakot: vol.7, no.11.

Az adatbázisban külön adatrekord tárolja a folyóirat alapadatait, és külön-külön rekordok az egyes folyóiratszámok nyilvántartási adatait.

A konferenciaanyagok előválogatását szintén belső adatbázis segíti. Az előzetes kiválasztás a *Meeting Agenda* nevű nyilvános adatbázisban (*Télésystemes-Questel* szolgáltatóközpont) futtatott újdonságfigyeléssel történik. Az így kiválasztott konferenciákat azonnal beviszik a belső konferencia-adatbázisba. Ebben a konferenciákat elsődlegesen a konferencia helye és pontos időpontja szerint tartják nyilván, mivel a konferencia neve elég gyakran megváltozik az első meghirdetés és a konferenciaanyag kiadása közötti időben. (Egybeesésnél másodlagos ismérvként a név, a téma, a rendező szerv neve segíti a megkülönböztetést.)

A folyóirat-adatbázishoz hasonlóan ezt az adatbázist is a kiadványnak indexelési kötegekre való szétszedéséig használják nyilvántartásra. A konferencia-adatbázist kezelő alrendszer körülbelül egy hónappal a konferencia meghirdetett kezdő időpontja előtt generál figyelmeztetést, ennek nyomán kéri meg a kiadványt a konferencia szervezőbizottságától.

A deskriptív katalogizáláshoz kapcsolódó online adatrögzítés során a konferenciák adatait — a folyóiratok adataihoz hasonló módon — ebből az adatbázisból viszik be a feldolgozott dokumentumok rekordjaiba, így itt sincs szükség ismételt bebillentyűzésre. A konferenciaanyagokat két szinten dolgozzák fel, éppúgy, mint a cikkgyűjteményeket: külön adatrekord készül a kiadvány egészéről és egy-egy külön rekord az egyes referált előadásokról.

## Válogatás, deskriptív katalogizálás, első adatrögzítés

A megkapott folyóiratszámokból és konferenciaanyagokból most már konkrétan ki kell jelölni a referálandó cikkeket, illetve előadásokat. A konferenciaanyagok egészéről, a könyvekről és az egyéb dokumentumokról el kell dönteni, hogy egyáltalán feldolgozandók-e. Végül a könyvekről és a terjedelmesebb kutatási jelentésekről el kell dönteni azt is, hogy fejezetenként is feldolgozzák-e, ha igen, akkor ki kell jelölni ezeket a fejezeteket.

Mindezt a kijelölő munkát az indexelő csoportok vezetői végzik. Egyes adatbázis-készítő, referáló folyóiratokat szerkesztő szervezetek a referálókra, indexelőkre hagyják rá a referálandó anyag kiválasztását. Ezzel azonban elkerülhetetlenül fő kiválasztási szemponttá lép elő, hogy az adott dokumentum mennyire könnyű vagy nehéz referálni, illetve indexelni. Az INSPEC nem engedi meg magának azt a luxust, hogy ilyen hamis kiválasztási szempont érvényre jusson. A csoportvezetők, akik a kiválasztást végzik, maguk nem indexelnek, a referálandó dokumentumok kiválasztásán kívül az indexelés lektorálása a feladatuk.

A kiválasztásnak egyébként a fő szempontja az, mint már mondtuk, hogy a dokumentum belesik-e az INSPEC témakörébe. A témába eső dokumentumok közül többnyire csak a kisebb jelentőségűeket hagyják ki: a hirdetéseket, olvasói leveleket, hibaigazításokat, híreket, szerkesztőségi üzeneteket stb.

Az indexelők 11 szakmai csoportot alkotnak, csoportonként 4–8 emberrel, összesen 45 indexelővel. A csoportok közül 5 foglalkozik a fizika különböző területeivel, 3 az elektromossággal és elektronikával, 1–1 a számítástechnikával, az irányítástechnikával és a távközléssel. A szakmai elhatárolás nem éles, egy adott folyóirat, konferenciaanyag, cikkgyűjtemény mindig egyetlen csoporthoz kerül, akkor is, ha a benne található dokumentumok egy része tulajdonképpen másik csoport szakterületéhez tartozna.

A feldolgozandó dokumentumokat általában a tartalomjegyzéken jelölik ki. A kijelöléskor kapják meg a dokumentumok azt az azonosítójukat (kézzel a tartalomjegyzékre írva), amely aztán a teljes feldolgozás alatt végigkíséri őket. Ez a jelzet több részből áll. Folyóiratcikknél például tartalmazza a folyóirat és a folyóiratszám azonosítóját (ez az évszám utolsó két jegye és az éven belüli szám három jegyre kiegészítve) és a számon belüli sorszámot. Az azonosító első két része egyébként a folyóirat-adatbázisban a folyóiratszámot leíró rekord azonosítója is, a folyóirat egészét leíró rekord azonosítója

pedig ennek az azonosítónak az első része, csupa nullából álló második résszel.

A kijelöléskor a tartalomjegyzékre ráírják a teljes paginációt is. (Ott eredetileg általában csak a kezdő oldalszám található.)

A referálandó dokumentumokat kijelölő csoportvezető jelöli ki az indexelési kötegeket is. A vékonyabb folyóiratszámok egyetlen köteget alkotnak, a vastagabb számokat, konferenciakiadványokat, cikkgyűjteményeket azonban tíz-tizenöt dokumentumnál többet nem tartalmazó kötegekre bontják. Ugyanekkor mondják meg, hogy az egyes kötegeket ki fogja indexelni, esetleg melyik külső munkatárs fogja kivonatosolni.

A kijelölés után kerül az anyag a deskriptív katalógizálókhoz, akik kisméretű, egyszerű munkalapra dolgoznak, hiszen csak néhány adatot kell felírniuk: az azonosító mellett a paginációt, a hivatkozások számát, a konferencia jelét akkor, ha folyóiratban jelent meg a konferenciaanyag. A többi adatot vagy közvetlenül a címlapról, illetve a cikk fejlécéről gépelik be az adatrögzítők, vagy házi adatbázisból veszik át. A már említett folyóiratadatokon és konferenciaadatokon kívül a szerzők munkahelyének adatait is adatbázisban tárolják, és ha már megvan, onnan viszik be a referátum rekordjába. Ha még nincs meg, akkor az első adatrögzítés során beviszik ebbe az intézmény-adatbázisba is.

A deskriptív katalógizálás után kerül sor az első adatrögzítésre. Ezt interaktív módon belső adatrögzítők végzik, akiket betanítanak a szerkesztést segítő számítógépes rendszer megfelelő részének használatára. Az interaktív adatrögzítés során másolják át a létrehozott adatrekordba a folyóirat-adatbázis, illetve a konferencia-adatbázis és az intézmény-adatbázis megfelelő adatmezőit. Az új adatok begépelését megfelelően kialakított képernyőformátumok segítik.

Ebben a műveletben, az újonnan begépelte és az átmásolt adatokból áll elő első formájában az az adatrekord, amely majd végső formájában a kérdéses dokumentum INSPEC adatbázisbeli rekordja lesz, közbülső állapotában pedig az elvégzett és a hátralevő munkafázisokat is nyilvántartja. Ebbe az adatrekordba kerül majd később az adott dokumentum leírására szolgáló minden további adatmező.

Az interaktív adatrögzítés után a következő éjszaka kötegelt üzemmódban korrekturalista készül a létrehozott adatrekordokból. Ezen a nyomtatott listán korrektúrázzák a bevitt adatokat ugyanazok, akik a deskriptív katalógizálást végezték. A felfedezett és bejelölt hibák korrigálását ismét terminálról, interaktívan végzik az adatrögzítők. Ebben az adatrögzítési és korrigálási munkamenetben szokásos billentyűzetű terminált használnak, így csak az álta-

lános használatú terminálokon megszokott karakterek, az EBCDIC-kódtáblázat karakterei használhatók. Az ezektől eltérő karaktereket táblázatból kikereshető karakterfüzérékkel helyettesítik (pl. H<sub>2</sub>O helyett H/sub 2/O).

Az adatrögzítés és a korrektúra után a folyóiratszámok, konferenciaanyagok, cikkgyűjtemények, fejezetenként is indexelt könyvek szétbontása következik indexelési kötegekre. Itt a szó szoros értelmében vett fizikai szétszedésről, széttépésről van szó. Nem másolnak, nem fotóznak az indexelő számára, de nem is őrzik meg könyvtárban használható módon a kötetet. A beszerzett dokumentumok a feldolgozás során megsemmisülnek, ugyanis az INSPEC-nek az eredeti dokumentumok beszerzésével semmi más célja nincs, mint feldolgozni azokat. Könyvtárat nem tart fenn, felhasználóinak az eredeti dokumentumokról másolatot nem szolgáltat. Amelyik felhasználó a dokumentumok eredetijéről másolatot kér, azt más intézményhez utasítják, legtöbbször a *British Library Lending Division*-hoz, konferenciaanyagokra vonatkozóan esetleg az *Institution of Electrical Engineers* könyvtárához. Így azután bármilyen másolás csak a költségeket és az átfutási időt növelné. (Az átfutási idő bibliográfiai adatbázisnál és referáló folyóiratnál a minőséget elsődlegesen meghatározó tényezők közé tartozik.)

A folyóiratot, konferenciakiadványt, cikkgyűjteményt annyi darabra tépik szét, ahány indexelési köteget jelölt ki benne a csoportvezető. Ezután minden cikkhez kitépik az első oldalát tartalmazó lapot, és ha nem ezen a lapon van (mint itt, a TMT-ben), akkor a kivonatot tartalmazó lapot is. Egyetlen esetben kényszerülnek gyorsmásolat készítésére, ha azonos lapon van két különböző cikk eleje vagy kivonata, esetleg az egyik eleje és a másik kivonata, mégpedig olyan elrendezésben, hogy ollóval nem különíthetők el jól.

Egy indexelési köteg tehát a következőkből áll: minden indexelendő dokumentum első oldala (rávezetve a szerzők nevét és az első szerző munkahelyét, ha az egyébként az utolsó oldalon volt), az esetleges külön lapon levő kivonata és az indexelési munkalapja, valamint a kötet vagy folyóirat megfelelő részének az említett lapok kitépése utáni maradék, dokumentumokra való szétszedés nélkül. Ez a maradék azonban csak akkor tartalmazza biztosan a cikkek teljes maradék szövegét, ha olyan folyóiratról, kötettről van szó, amely minden cikket, előadást, fejezetet következetesen páratlan oldal tetején kezd. Az így kialakított indexelési köteget megfelelő méretű fóliazacskóba helyezve adják át további feldolgozásra. Ebben a zacskóban folytatja azután az útját az összes további feldolgozási lépésen át a megsemmisítésig.

## Szerkesztési-nyilvántartási rendszer, adatállományok

Az INSPEC számítógépes szerkesztési-nyilvántartási rendszerének fő elve, hogy a szerkesztést és a szerkesztési tevékenységek nyilvántartását egyetlen közös rendszerben végzik. Ez a közös rendszer nappal interaktív üzemmódban fut az INSPEC kisebbik, 2 Mbájt központi tárral rendelkező VAX 11/750-es gépén és a vele hálózatra kapcsolt terminálokra, éjjel pedig kötegelt üzemmódban a két VAX 11/750-es gépen együtt. A programrendszert az RMS adatbázis-kezelő rendszer felhasználásával írták, BASIC programrészekkel kiegészítve; másfél programozó mintegy kétévi munkájával készült. A rendszer másik lényeges elve, hogy egy adat csak egyszer kerüljön adatrögzítésre. Amelyik adat már bent van valamelyik adatállományban, az onnan felhasználható legyen más adatállományokhoz is.

A számítógépes rendszer középpontjában a szerkesztési adatállomány áll. Ebben minden feldolgozott dokumentumnak egy rekordja van a deskriptív katalogizálást követő első adatrögzítéstől az archiválás utáni törlésig. Ez az egy rekord szolgál a végleges adatbázisrekord felépítésére is és a dokumentum sorsának nyilvántartására is. Ez a rekord eleinte csak a deskriptív katalogizálás adatait és a nyilvántartó adatmezőket tartalmazza, a többi adatalem a fő adatrögzítési lépés nyomán kerül bele. A nyilvántartó adatmezőbe belekerül minden munkamozzanat elvégzésének a ténye és időpontja. Így a munka időbeli lefutása is nyomon követhető, az esetleges fennakadások felfedhetők. Ez nagyon fontos az átfutási idő, a referálási késés minimumra szorítása végett. A "gyalogos" munkamozzanatok elvégzését interaktívan, terminálról írják be a rekordba (erre menürendszerű nyilvántartó alrendszer van), a programmal végzetekét a feldolgozó program írja be automatikusan. A programmal végzett munkamozzanatokhoz a nyilvántartó mezőket is a programból használják fel. A program vizsgálja meg, hogy a rekord készültségi állapota megfelelő-e egy nyomtatott kiadványba vagy adatbázis-szekcióba való fölvételre, illetve hogy ennek a kiadványnak vagy szekciónak az előző számába nem került-e már bele, hogy az összes termékbe bekerült-e már, ami az archiválás és az ezt követő törlés feltétele stb.

Ugyanez az adatbázis alkalmas arra, hogy a megfelelő jelzőkaraktereket viselő rekordok alapján bizonyos összesített adatokat megkapjanak belőle, például az adott hónapban indexelt dokumentumok számát (indexelőnként egyénileg is), vagy az INSPEC-A szekció következő számának szerkesztéséhez az adott pillanatban készen álló rekordok számát.

A szerkesztési állományba rendszeresen bekerülő adatokat a folyóirat-adatbázisban, a konferencia-adatbázisban, valamint az intézmény-adatbázisban tárolják és karbantartják. Ezek közül az első kettő egyben az előkészítő munkák, főleg a beszerzés nyilvántartására is szolgál.

A szerkesztés során ellenőrzésre alkalmas a teaurusz-adatállomány és az osztályozási rendszer adatállománya, amelyek egyben ennek a két dokumentumnak a fejlesztésére is szolgálnak, és a helyesírást ellenőrző szótár, amelyről a korrektúra kapcsán szólunk részletesebben.

A rendszer további nagyobb adatállományai a különböző kiadványok szerkesztésének közbülső állapotát tárolják.

Végül nagy jelentősége van az archív állományoknak. A rendszerbe bevitt rekordok egyetlen archiválási eszköze ez az állomány. A rekordok ebbe akkor kerülnek át a szerkesztési adatbázisból, ha már minden kiadvány kiválogatási folyamatán átmentek. Egyetlen egységes archív állomány van, ezért ennek a rekordjaiba bekerül minden olyan adatmező, amelyeknek bármely kiadvány készítésekor szerepe van. Ebből tehát szükség esetén bármelyik kiadvány bármelyik száma feltámasztható. Rendszeresen azonban csak az INSPEC adatbázis régebbi számai mágnesszalagjának az előállítására használják, ha egy régebbi felhasználó egy megrongálódott szalagjának a pótlását kéri, vagy egy új előfizető nem a pillanatnyilag aktuális szalagnál akarja kezdeni az előfizetést, hanem egy előző időszak anyagát is kéri. Az aktuális mágnesszalagokat csak annyi példányban készítik el, ahányra előfizetés van, ebből tárolható tartalék nincs. Az archív adatállomány nem tagozódik sem havi számokra vagy évfolyamokra, sem szekciókra; egyetlen folyamatos soros, mágnesszalagos állomány, amelyből szükség esetén a rekordazonosítók és a szekciójelzetet is tartalmazó referátumazonosítók alapján lehet kiválogatni a kért korábbi számba sorolt rekordokat.

## Kivonatolás és indexelés

A csoportvezetők kijelölése alapján elkészült indexelési kötegek többsége azonnal az indexelőkhöz kerül. Ez történik minden angol nyelvű dokumentummal, valamint a francia és német nyelvű dokumentumok többségével. Az utóbbiak kisebb része, valamint a többi nem angol nyelvű dokumentum a mintegy 100 külső munkatárs közül valakihez kerül. Ezek a külső munkatársak készítik el a dokumentum angol nyelvű kivonatát, illetve hagyják jóvá, egészítik ki, korrigálják nyelvi szempontból az eredeti dokumentumhoz készített angol nyelvű ki-

vonatot. Ugyancsak ők készítik el a dokumentum címének angol fordítását és esetleges kiegészítését. Egyedül a japán anyagokhoz nem találnak elég szakembert, akire a kivonat megírását rábízhatnák. Ezért a japán nyelvű dokumentumokat általában lefordítatják, részben saját külső fordítókkal, részben fordítóirodával. Ezeket azután fordítás alapján a belső indexelők kivonatolják.

Az indexelést szinte kizárólag belső munkatársak végzik, a 45 folyamatosan dolgozó belső indexelő mellett mindössze 4–5 alkalmilag dolgozó külső indexelőjük van. A nem angol nyelvű dokumentumokat a kivonat és a címfordítás alapján indexelik. Kivétel ez alól a francia és német dokumentumok jelentős része, mert az indexelők között vannak olyanok, akik ezen a két nyelven megfelelő szinten értenek.

Az indexelők mindig annak a területnek a szakemberei, amelynek az irodalmával foglalkoznak: fizikusok, villamosmérnökök, számítástechnikai szakemberek stb. Kivétel nélkül mind olyan kezdő szakemberek, akiknek az INSPEC az első munkahelyük.

Fizikusok, villamosmérnökök, számítástechnikai szakemberek számára nem igazán vonzó munka az indexelés, pedig nélkülük az adatbázis nem hozható létre. Éppen ezért az intézmény vezetőinek nagy gondot okoz, hogy ezeket a fiatal szakembereket mivel motiválják. Ezt a gondot igazán még nem sikerült megoldaniuk. Ennek ellenére a fluktuáció itt nem olyan nagy, hogy a munkát veszélyeztetné: a 45 indexelő közül évente átlag 4–5 cserélődik, vagyis egy-egy indexelő átlagosan tíz évet tölt el ezen a helyen. Az új indexelők betanítására külön kézikönyvet írtak.

Az indexelők teljesítménye közel 100 dokumentum indexelése hetenként. Mivel munkaidejüknek körülbelül a felében foglalkoznak ténylegesen indexeléssel, ez azt jelenti, hogy ilyenkor mintegy 5 dokumentumot indexelnek óránként. Ezenkívül, amelyik dokumentumból szükséges, abból munkaidejük másik felében kivonatot készítenek, vagy a már meglévő kivonatot ellenőrzik, esetleg kiegészítik, s ha azt nem angol anyanyelvű írta, akkor nyelvíleg is javítják.

Az indexeléskor – éppúgy, mint a deskriptív katalogizálásakor – kizárólag kézzel írnak. Az indexelés és az adatrögzítés között nem gépelik le, amit leírtak, hanem részben kézírással, részben a dokumentum első oldalán és az (esetleg szintén kézzel írt) kivonatban aláhúzással kijelölve kerül az anyag az adatrögzítőkhöz. Ezért a munkatársak kiválasztásánál lényeges szempont az olvasható kézírással. A jelentkezők közül elég nagy arányban kénytele-

nek visszautasítani olyanokat, akik ennek a kritériumnak nem felelnek meg.

Az indexelők kisméretű, egyszerű szerkezetű munkalapon dolgoznak, mivel a bibliográfiai adatokat nem kell az adatlapra írniuk, azok már az adatbázisban vannak, az indexelés jelentős része pedig a dokumentum első oldalán kijelöléssel történik.

A cím fordítása vagy kiegészítése nem erre az indexelési munkalapról kerül, hanem a deskriptív katalogizálás munkalapjára. A cím adatrögzítési utasításában szerepel, hogy az adatrögzítőnek a cím első betűjét kivéve minden nagybetű helyett kisbetűt kell gépelnie. Így az indexelőnek csak azt kell bejelölnie (dupla kék aláhúzással), ahol mégis meg kell hagyni a nagybetűket (nevek, rövidítések, kémiai képletek és vegyjelek, mértékegységek stb.).

A kivonatot nem a munkalapról írják, hanem külön papírra. Mód van arra is, hogy a meglévő kivonatot elfogadva, azt külön papíron folytassák, kiegészítsék.

A dokumentum eredeti nyelve a munkalapon az első mező. Ha a dokumentum fő szövege több nyelvű, akkor több nyelv is bejelölhető (ez természetesen nagyon ritka). A kivonat nyelve nem indexelhető. Ha viszont egynyelvű a szöveg, és ez a nyelv az angol, akkor a nyelvmező üresen marad. Ez csökkenti az indexelő munkáját, mivel a cikkek 84%-a ilyen.

A következő mező a munkalapon az osztályozási jelzeteké. Az *A* szekció osztályozási jelzetei számára 7, a *B*, *C* és *D* szekciók jelzetei számára 4–4 hely van, de szükség esetén az előre nyomtatott szekciójelzet átírható. Az összesen 19 beírható jelzet gyakorlatilag mindig elegendő, bár elvileg az adható jelzetek számának nincs felső határa. A több szekcióba is besorolt dokumentumnak viszont minden szekciója számára legalább egy-egy jelzetet kell kapnia.

A deskriptorok és a nyomtatott referáló folyóirat tárgymutató-kategóriái számára közös mező van az adatlapon. Ezeket ugyanis párban kell adni; minden deskriptorhoz tartozik egy tárgymutató-kategória. Az így adott kategóriák szerint kerül majd be a dokumentum a nyomtatott folyóiratok tárgymutatójába. Tehát a tárgymutató nem utólag készül, mint a hagyományos szerkesztési módszernél, hanem már az indexelés során. A tárgymutató készítésének ezután következő lépései már teljesen gépesítve vannak (lásd lejjebb). A tárgymutató-kategóriákat az indexelő köteles a korábbi évfolyamok tárgymutatójából venni, ettől csak nagyon indokolt esetben térhet el, például lényeges újdonság, újonnan keletkezett fogalom esetén. A munkalapon 6 deskriptor–tárgymutató-kategória pár számára van hely, de pótlapra továbbiak is írhatók. A tárgymutató-

kategória gyakran megegyezik a deskriptorokkal, akkor nem kell újra kiírni.

Minden dokumentum a tárgymutató számára kap egy néhány szavas tartalmi címet is, amely feltétlenül a cikk legfontosabb koncepciójával kezdődik, és a dokumentumot a tárgymutatóban a megfelelő kategóriában képviseli. Ez kivételesen lehet az eredeti cím is, ha az olyan, hogy erre a célra minden szempontból alkalmas. Ilyenkor nem kell a munkalapra ismételtelen leírni.

A szabad tárgyszavakat a címben, a kivonatban és az első oldal szövegében elsősorban aláhúzással jelölik ki. Ezen túl a munkalapra is írhatók további tárgyszavak, sőt, pótlapra is lehet még továbbiakat írni. A címben és a kivonatban aláhúzással jelölt szabad tárgyszavakat az adatrögzítő különleges jelek (dupla szögletes zárójelek) közé téve gépeli le. Így azokat a feldolgozó program automatikusan kiválogatja a szövegből, nem kell őket újra leírni. Külön szabad tárgyszóként tehát az adatrögzítő csak a munkalapra vagy a pótlapra írt és a főszöveg elején kijelölt kifejezéseket billentyűzi be.

A tárgyalásmód (kísérleti, elméleti stb.) jele előre nyomtatva szerepel a munkalapon. Ezt az indexelőnek csak bekarikázással kell jelölnie. (Több jelzetet is bekarikázhat.)

Az indexelést és a kivonatot a szerkesztési csoportok vezetői lektorálják.

### Adatrögzítés, korrektúra, felszabadítás

A kivonat és az indexelés lektorálása után következik a második adatrögzítés. Az ekkor beírandó szöveg sokszorosa az első adatrögzítésnél leírtaknak. Ezt a nagyon időigényes munkát három külső szolgáltatóvállalattal végeztetik. A szerződés szerint mindhárom vállalattól tíz napon belül kell visszaérkeznie a mágnesszalagra írt anyagnak.

Ennél az adatrögzítésnél már nem elégszenek meg az EBCDIC-karakterkészlettel, mint az első adatrögzítés során. Itt már szintjelek (shift) segítségével négy karakterkészletben összesen 471 karaktert használnak, mindegyiket 8 különböző karaktertípusban. A 471 karakter között számos matematikai jel található, görög betűk, alsó és felső indexek, pénznemek jelei, különleges jelek stb. A nyolc karaktertípus között a normál, a kurzív és a félkövér betűk mellett a szokásosnál kisebb és nagyobb betűméretek (címbetűk) is vannak.

Az adatrögzítésről érkező adatrekordokat először kötegelte üzem módban összefuttatják a rekordok már korábban rögzített részével. Ezután futtatják — ugyancsak kötegelten — a hibaellenőrző programot. Ennek az ellenőrző funkciói között a legtöbb olyan,

ami más szövegfeldolgozó rendszerekben is megszokott. Ilyenek például egyes mezők ellenőrzése adattípusra (numerikus, alfanumerikus stb.), az osztályozási jelzetek ellenőrzése ellenőrző karakterrel és a gépben tárolt osztályozási rendszerrel való összevetéssel, a deskriptorok ellenőrzése a gépben tárolt tezaurusszal, a zárójelek párosságának ellenőrzése, annak az ellenőrzése, hogy egyes mezők (kivonat, szerzők neve, szerző munkahelye stb.) nagybetűvel kezdődjenek, hogy a szerzők keresztneveinek a kezdőbetűi nagybetűk legyenek.

A hagyományos ellenőrzési módok mellett van a programnak egy újdonságnak tekinthető ellenőrzési módja is. Ez pedig az ellenőrző szótárral való összehasonlítás. Az ellenőrző szótárral a másképp nem ellenőrizhető mezők (cím, kivonat, szabad tárgyszavak, szerzők stb.) minden egyes szavát összevetik. Ha egy szó a szótár egyetlen szavával sem egyezik, a program figyelmeztető üzenetet generál. Ennek a feldolgozása során három dolog történhet. Ha valóban hibás a megjelölt szó, kijavítják a korrektúra során. Ha a figyelmeztető üzenet abból ered, hogy a szótárban van hiba, akkor ott javítanak. Végül ha a figyelmeztető üzenet oka egy új, a szótárban még nem szereplő szó, akkor a kérdéses szóval kiegészítik a szótárat. A szótár induló változatát az archív álmány korábbi egyéves szakaszából programmal hozták létre. 1984 februárjában, az induláskor a szótár mintegy 90 000 szót tartalmazott. Ez a figyelmeztető üzenetek feldolgozása során bevitt új szavakkal 1985 szeptemberére már több mint a kétszeresére, 187 435 szóra gyarapodott. Még így is kevesebb mint 10 Mbájt lemezterületet foglal el. A szótár leggyakoribb 10 000 szava állandóan a belső tárból van a futás gyorsítása céljából.

Az ellenőrző program futása után a rekordokat printer-plotteren nyomtatják ki. Ez mátrixnyomtató módjára nyomtatja az EBCDIC-karaktereket, a többit pedig plotter módjára rajzolja. Így végül is a 471 karakterből álló készlet minden karaktere kiírható, összességében viszonylag gyorsan, hiszen a nyomtatandó karakterek túlnyomó részét teszik ki a gyorsan nyomtatható EBCDIC-karakterek.

Ezen az ellenőrző listán minden rekord után hibaüzenetek és figyelmeztető üzenetek listája sorolja fel a felfedezett vagy feltételezhető hibákat. Az ellenőrző szótárral kiszűrt szavakat ezenkívül a szó elé írt feltűnő jel (csúcsán álló tömör négyzet) is mutatja.

Az éjszakai műszakban nyomtatott ellenőrző listákat reggel a korrektorok kapják meg. Ők a hibajelzések és figyelmeztető üzenetek alapján bejelölik a szükséges javításokat. Ezután hagyományos módon végigolvasva is ellenőrzik a szöveget. Így fedezik fel a géppel — a mesterséges intelligencia mai fejlett-

ségi szintjén — még felismerhetetlen hibákat, például a szórendi hibákat, a vesszőhibákat, bizonyosfajta szóközhibákat.

Néhány indexelési köteg papíron való feldolgozása után ülnek a korrektorok a terminálhoz. Ez különleges, az INSPEC céljaira tervezett intelligens terminál, amelyen a teljes karakterkészletnek mind a 471 karaktere mind a 8 betűtípusban megjeleníthető és bebillentyűzhető. Ennek a képernyőjére hívják elő a javítandó rekordot, majd a terminál (tulajdonképpen mikroszámítógép) autonóm üzemmódban működő szövegszerkesztő programjával elvégzik a papíron már előzőleg bejelölt javításokat. Ezután — már a hálózat közvetítésével a nagy gépen — újra lefuttatják erre a rekordra a hibaellenőrző programot, amely nemcsak kötegelte üzemmódban működhet, hanem terminálról párbeszédés üzemmódban is elindítható. Ha netán még mindig hibát találnak, akkor újabb javítás és ellenőrzés következik. Csak a terminálon már hibátlanak ítélt rekordot írják vissza kijavított formájában az adatbázisba.

A korrekció után következik a feldolgozás utolsó manuális lépése: a kész rekordok felszabadítása. Erre csak azoknál a rekordoknál ad módot a program, amelyekben a legutolsó gépi ellenőrzés nem talált hibát. A felszabadítást rendszerint indexelési kötegenként végzik, de szükség esetén rekordonként külön-külön is elvégezhető. A felszabadítás azt jelenti, hogy a rekordot késznek, hibátlanak, az adatbázisban, a nyomtatott kiadványokban és az egyéb szolgáltatásokban felhasználhatónak minősítik, a programnak engedélyezik a további, már közvetlen emberi beavatkozás nélküli, automatikus lépéseket. Ettől kezdve a rekord sorsát a benne található jelzőkarakterek (flagék) és egyéb mezőtartalmak (például a szekciójelzetek) irányítják. A jelzőkarakterekből ismerik fel az egyes feldolgozó programok, hogy a rekordot felhasználhatják, és ezeken jelölik meg ugyanezek a programok, hogy a rekordot bizonyos végtermékbe már beépítették. A feldolgozó programok egyébként a közbülső munkafázisokban nem ezeket a teljes adatrekordokat mozgatják, hanem csak az úgynevezett triggerállományokat, amelyek a rekordazonosító mellett rendezési kulcsokból és jelzőkarakterekből állnak.

Az INSPEC-végtermékek tartalmilag három csoportba oszthatók: szignaletikus termékek, referáló termékek és szelektív információterjesztés. Fiktív negyedik végterméknek tekinthető az archív állomány. Formailag a szignaletikus és referáló termékek nyomtatottak vagy mágnesszalagosak lehetnek. (A mikroformátumban eladott termékek az INSPEC-nél csak a nyomtatott termékek fénymásolatai.) Saját online szolgáltatást az INSPEC nem

folytat, viszont sok online szolgáltatóközponttal van szolgáltatási szerződése. A szelektív információterjesztési formák közül a szabvány profilos változatok nyomdai úton vagy sornyomatatóval készülhetnek. Az egyedi profilokkal folytatott szelektív információterjesztés mindig sornyomatós.

## Szignaletikus és referáló termékek

A szignaletikus (current awareness) termékek kivonat nélkül készülnek, de a megjelenésük gyorsabb, mint a referáló termékeké. Ide sorolhatók a nyomtatott *Current Papers in Physics*, *Current Papers in Electrical and Electronics Engineering* és *Current Papers on Computers and Control* folyóiratok, valamint az *INSPEC-2* mágnesszalag. Az *INSPEC-2* havonta kétszer jelenik meg, és nem oszlik szekciókra. A *Current Papers* folyóiratok közül a fizika havonta kétszer, a másik kettő havonta egyszer jelenik meg.

A referáló termékek a Science Abstracts-család már korábban említett négy folyóirata, valamint az igazi INSPEC adatbázis, az *INSPEC-1* mágnesszalag. Az *INSPEC-1* a referáló folyóiratoknak megfelelő négy szekcióra oszlik, és szekciónként is, egészként is előfizethető. A *Physics Abstracts* havonta kétszer, a többi három folyóirat havonta egyszer jelenik meg. Az *INSPEC-1* adatbázis havonta kétszer készül, de ebből csak az egyik kiadásban található a tisztán *B* szekciójú (elektromosság és elektronika), csak a másikban a tisztán *C* szekciójú (számítástechnika és irányítás) és a tisztán *D* szekciójú (irodai információs technika és telematika) rekordok. Az *A* szekcióba (fizika) sorolt és a többi szekcióba is besorolt rekordok bármelyik kiadásba bekerülhetnek.

A nyomtatott termékek készítésének első lépése az adott termékben felhasználható rekordok kiválogatása és témák szerinti (osztályozási kódok szerinti) rendezése. Ekkor kapja meg a referátum az azonosítóját is, amelynek viszont be kell kerülnie az adatbázisba, tehát az adatbázis csak a megfelelő nyomtatott termék rekordjainak kiválogatása után állítható elő. Ezután az egyik programon a fényszedésre való előkészítés következik a folyóiratjegyzékkel kapcsolatban már említett programmal. Az eltérés annyi, hogy itt a program előre paraméterezett változata használható az oldalak és hasábok, a rekordok (bekezdések) kialakítására, a betűtípusok kijelölésére, a fejlécek kialakítására stb. Másik két programág szolgál a tárgymutató és a szerzők mutatója kialakítására, de ezek az ágak csak félévenként futnak végig. További programok készítik a félévenkénti mutatókból a négyévenkénti kumulált indexeket.

A mágnesszalagok tulajdonképpen minden előfizető számára külön-külön, testre szabottan, az egyéni kívánásoknak megfelelően készülnek. Ez gépidőben nem jelent lényeges többletet, hiszen mágnesszalagot másolni úgyis csak egyenként lehet, valódi sokszorosításra nincs mód. A felhasználó megválaszthatja a karakterkészletet (csak nagybetűk BCD-kódban vagy nagy- és kisbetűk EBCDIC-kódban, vagy a nyomtatott kiadványoknak megfelelő teljes karakterkészlet 471-féle karakterrel), a sávok számát (7 vagy 9), az írássűrűséget (800 vagy 1600 bpi), az adatbázistípust (INSPEC-1 vagy INSPEC-2, tehát referáló vagy szignaletikus), az INSPEC-1 esetén a kért szekciókat. A sok választási lehetőség együttesen nagyon sok szalagváltozatot eredményez. Ezért külön nyilvántartási rendszert alakítottak ki a mágnesszalag-készítés és -küldés adminisztrálására. Ez a rendszer az egyik mikroszámítógépen autonóm üzemmódban fut, a fő rendszertől teljesen függetlenül. Minden egyes készülő mágnesszalagra megadja a szükséges adatokat, az elkészítéséhez szükséges programok felsorolását, és két öntapadós címkét is nyomtat hozzá három-három példányban. Az egyik címke a mágnesszalag adatait tartalmazza, a másik a felhasználó postai címét. A három példány közül egy a mágnesszalagra, illetve a csomagolásra kerül, a másik a külön levélben küldött értesítésre, amely egyben visszajelző űrlap a küldemény megérkezéséről és hibátlan vagy hibás voltáról, a harmadik postai cím ennek az értesítésnek a borítékjára, a harmadik mágnesszalagcímke pedig a gépteremnek kell.

A mágnesszalagokat előállító programok részben még az öreg ICL-gépen futnak, a 7 sávós és a 800 bpi írássűrűségű szalagot előállító programágak távolatlag is ott maradnak, az új számítógépek nincsenek felszerelve a már elavultnak tekinthető mágnesszalagegységekkel. Mindkét gépen minden közbülső lépésben lemezes adatállományokat állítanak elő, amelyeknek a rekordjai csak rendezési kulcsokból és azonosítókból állnak, csak az egyik gépről a másikra való átadásban és a végső másolási lépésben írják a teljes rekordokból álló adatállományt mágnesszalagra.

### Szelektív információterjesztés

Az INSPEC — az adatbázis-építés gazdaságosságát javítandó — szelektív információterjesztést is folytat, közvetlenül a belső adatállományból. Ez részben egyéni profilok szerint, részben szabványos (standard) profilok szerint történik (vö. a hazai témafigyeléssel).

A szabványos profilok szerinti szelektív információterjesztés egyik típusa nyomtatott havi kiadvány *Key Abstracts* címmel. Nyolc ilyen kiadványuk van. Ezek egy-egy nagyobb és kiemelkedően népszerű témakört referálnak (pl. félvezető eszközök, elektronikus áramkörök, mérések és műszerek a fizikában), de együttesen is csak kis részét fedik le az INSPEC teljes területének. Az egyes kiadványok anyagát egyrészt az osztályozási rendszer szerint választják ki, másrészt még ezen belül is kiválasztják a folyóiratok és konferenciakiadványok egy viszonylag szűk körét, így csak a legjelentősebb folyóiratok cikkeit és konferenciák előadásait referálják. Ez is indokolja a "Key Abstracts" címet, vagyis azt, hogy kulcsfontosságúnak tekinthetők azok a dokumentumok, amelyek kivonata szerepel ezekben a kiadványokban. Egy Key Abstracts-füzet terjedelme soha nem haladhatja meg a 24 oldalt.

A szabványos profilok szerinti szelektív információterjesztés másik típusa *Topics* címen a számítógép sornymotatóján, kártyaformában készül, hetenként. Ilyen szabványprofil 72 van. Ezek lényegesen szűkebb profilok, amint a Key Abstracts kiadványokéi, és gyakrabban is indítanak újat. A kártyák nem tartalmaznak teljes kivonatot a dokumentumokról, hanem csak a kivonat elejéből annyit, amennyi a bibliográfiai adatok mellett a kártyára ráfér. A dokumentumokat kizárólag az osztályozási jelzetek szerint választják ki.

Mind a Key Abstracts, mind a Topics keretében fontos kérdés a figyelt témakörök kiválasztása, amihez a témakör szakmai konferenciáit használják fel. Az INSPEC illetékes munkatársa a fizikusok, villamosmérnökök, számítástechnikai és irányítás-technikai szakemberek nemzetközi konferenciáinak résztvevőlistája alapján igyekszik előfizetőket gyűjteni a konferencia témakörének megfelelő új szabványos profil számára, a Topicshoz szűk témakörre szakosodott kisebb konferenciák, a Key Abstracts-hoz nagy kongresszusok résztvevőitől. Ha kellő számú előfizetést tud szervezni, akkor elindítják az új témát.

A szabványos profil szerinti szelektív információterjesztés harmadik típusa a *Japan Update*. Ennek hat szakmai sorozata van, és mindegyik kizárólag a japán folyóiratok anyagát tartalmazza. A hat téma túlnyomó részben lefedi az elektronika, a számítástechnika és az irányítás-technika területét, tehát mindazokat a területeket, ahol a japán szakirodalomnak elsődleges jelentősége van. A Japan Update is számítógépen nyomtatva, kártyaformában jelenik meg, de ebben a Topicsszal ellentétben folytatáskártyákat is készítenek, így teljes kivonatot ad.

Kártyaformában (15,24 cm · 10,16 cm-es, nagyjából levelezőlap méretű kártyákon), hetenként

kapják az eredményt az egyéni profil szerinti szelektív információterjesztés előfizetői is. A profilban osztályozási jelzetek, szabad tárgyszavak, tárgyalásmód, szerzők, nyelv, dokumentumtípus, folyóiratnév szerepelhet. Deszkriptorokra nem keresnek, ezért kénytelenek az indexelés során egyes deszkriptorokat megismételni a szabad tárgyszavak között. A cím és a kivonat kifejezéseire sem lehet keresni. A szabad tárgyszavaknál viszont lehetőség van nemcsak jobb, hanem bal oldali csonkolásra is. A keresés kötegetelt üzemmódú. A felhasználó választhat, hogy minden dokumentumról csak az első kártyát kéri-e (bibliográfiai adatok és a kivonat kezdete); vagy az első folytatáskártyát is a kivonat nagyobb részével, rövid kivonatok esetén a teljes kivonattal; vagy az összes folytatáskártyát, hosszú kivonat esetén is a teljes kivonattal; vagy végül a bibliográfiai adatok mellett kivonat helyett a deszkriptorokat. Választása tükröződik az előfizetési díjában. A keresőprogram a találatokon megjelöli azt a kifejezést, amely a találatot hozta.

Az egyéni profilokat írásbeli kérés és telefonon történő megbeszélés alapján állítják össze. A felhasználó a profilját tetszőleges sűrűséggel díjtalanul módosíthatja. Az eredményhez visszajelző nyomtatványt mellékelnek, amelyen a felhasználó egyenként megjelölheti a találatok relevanciáját. Ezt azonban kevés felhasználó küldi vissza, inkább levélben vagy telefonon jelzik időnként, hogy milyen irányban kéri a profil javítását vagy módosítását. Általában 70% körüli pontossággal meg vannak elégedve. Az egyéni profilokhoz az INSPEC-nek csak egy profilszerkesztője van, de szakmai konzultánsként mellette állnak az indexelők.

## Piac- és árpolitika

Az INSPEC bevételének legnagyobb részét még ma is a nyomtatott kiadványok, ezen belül is főleg a Science Abstracts előfizetése adja. Ezen a területen azonban az offenzív piacpolitikának már nincs kellő talaja. Az előfizetők száma hol lassabban, hol gyorsabban, de évről évre folyamatosan csökken. Különösen erős volt ez a tendencia a legutóbbi néhány évben, de még ahol nominálértékben megmaradt a nyomtatott kiadványokból származó bevétel, a reálértékben ott is kevesebb lett. Ha emeli is az INSPEC a nyomtatott kiadványok előfizetési díját, olyan arányban nem teheti, hogy ez kompenzálja az előfizetési létszám csökkenését, mert az ilyen mérvű drágulás az előfizetés csökkenését nagyon meggyorsítaná, tehát az összbevétel ettől nem nőne. Így azután a költségek egyre nagyobb és nagyobb részét hárítják át a számítógépes felhasználókra, az adatbázis ára, illetve a nyilvános szolgálta-

tóközpontokban a használat alapján kért jogdíja gyorsabban nő, mint a nyomtatott kiadványok ára. Ezt az utóbbi években megtehették, mert más adatbázisok ára is nőtt, sőt általában gyorsabban, mint az INSPEC-é. A több nyomtatott kiadványt együttesen előfizető felhasználók árkedvezményt kapnak.

A számítógépes felhasználás díjából a legnagyobb tétel az, ami a nagy nyilvános online szolgáltatóktól folyik be. Hét nemzetközi és három országos nyilvános online szolgáltató "pörgeti" üzemszerűen az INSPEC adatbázist. Velük az INSPEC egyedi feltételek szerint köt szerződést. A fizetendő ár alapja minden esetben egyrészt az online óradíjból, másrészt a kinyomtatott vagy megjelenített találatok nyomtatási formátumtól függő darabdíjából álló jogdíj. Ezt a szolgáltató ugyanazzal a programmal tartja nyilván, amellyel ő maga a felhasználói számláit készíti, így a díjak nagyon pontosan és eléggé megbízhatóan számolhatók el.

A nyilvános online szolgáltatásban a személyi számítógépek széles körű elterjedése és terminál gyanánt való egyre általánosabb használata óta kényes probléma az adatbázis rekordjainak saját számítógépes háttértárba való tömeges letöltése (downloading). Az INSPEC az egyik legelső adatbázis-előállító volt, amely felismerte, hogy aminek a tilalma ellenőrizhetetlen, azt nincs értelme megtiltani. Ezért elsőik között engedélyezte a korlátozott letöltést. Ebben három kategóriát állított fel. A legfeljebb 28 napos megőrzésre való letöltést legfeljebb 25-szöri kinyomtatással díjtalan, ha a kinyomtatott példányokból csak egyet adnak át külső felhasználónak. A letöltést végző intézményben belső felhasználásra való tartós megőrzésért, többszöri kinyomtatásért bevallás alapján, árjegyzék szerinti mérsékelt díjat kérnek. A letöltött rekordokból való nyilvános szolgáltatást, továbbeladást (ami már nehezen tartható teljesen titokban) csak egyedi szerződés szerint, jelentős díjért engedélyezik.

A tíz nagy online szolgáltatón kívül a többi felhasználó árjegyzék szerinti árat fizet az adatbázisért. Ez két részből áll. A mágnesszalagon kapott adatbázisért előfizetési díjat kell fizetni. Az előfizetési díj szekciónként változik. A több szekcióra is előfizető felhasználók a nyomtatott kiadványok előfizetőihez hasonlóan kedvezményt kapnak. Az előfizetési díj magában foglalja az első száz szelektív információterjesztési profil és az első ötven óra saját gépen való online keresés jogdíját. Az ár másik része az a jogdíj, amelyet a felhasználó a további szelektív információterjesztési profilokért és a saját számítógépen való további online keresésért évenkénti bevallás alapján fizet.

Az INSPEC saját szelektív információterjesztésének a bevételek szempontjából csak kiegészítő sze-



repe van. A nyomtatott kiadványok mikroformátumú változatából származó bevétel elhanyagolható. Ezekért egyébként ugyanannyit kérnek, mint a megfelelő nyomtatott változatért. A korábbi évek kiadványai lényegesen olcsóbbak, mint a frissek.

## Következtetések

Bár az INSPEC a világ adatbázis-előállítói között nagyüzemnek számít, mégsem tér el nagyságrenddel a hazai referáló folyóiratok előállításában megszokott mérettől. Ha egy az egyben nem is másolható le nálunk az INSPEC adatbázis-előállítási rendszere, módszereik sok részlete átvehető, vagy a hazai viszonyokra adaptálható.

Számítógépes szerkesztési és nyilvántartási rendszerük talán az első pillantásra egyeseknek számunkra elérhetetlennek tűnik. Ha azonban alaposabban végiggondoljuk, akkor kitűnik, hogy nem az. Két megamini-kategóriájú és néhány mikrokategóriájú számítógép már hazai viszonylatban sem számít nagyon nagy gépi kapacitásnak. A két VAX gép 2 Mbajt, illetve 3 1/4 Mbajt belső tára nálunk ma még meglehetősen nagynek számít, de már nem elérhetetlen, és néhány éven belül megszokott méret lesz. Ugyanez a helyzet az összesen 1755 Mbajt lemezes háttértárral. Amelyik periféria nálunk még különlegességnek számít, az nem is nélkülözhetetlen a rendszerben. Végül a lokális hálózati hardver és szoftver, illetve az adatbázis-kezelő szoftver megvásárolható, a 3 emberév saját prog-

ramfejlesztés pedig kimondottan kis munkamennyiségnek tekinthető.

Tehát néhány évre tervezett fejlesztésben már az INSPEC egész számítógépes rendszerének az adaptálása megcélózható nálunk is, a rövidebb távú, meglévő eszközökre alapozandó fejlesztésekhez pedig kiválaszthatók az INSPEC rendszeréből azok a részletmegoldások, amelyek a mai eszközállományra alapozva jelentős segítséget adnak az adatbázis-építéshez, és fontos funkciókat látnak el a számítógéppel segített szerkesztés feladataiból.

Végkövetkeztetésként azt állapíthatjuk meg, hogy aki Magyarországon adatbázis-építésbe fog, annak minden részletében hasznos az INSPEC rendszerének a megismerése.

Befejezésül szeretném köszönetemet kifejezni az INSPEC munkatársainak, akik lehetővé tették, hogy az INSPEC feldolgozási rendszerét, munkafolyamatát megismerjük. Elsősorban *T. M. Aitchison* igazgatónak; *P. C. Clague* igazgatóhelyettesnek; *G. E. Mears*-nek, a marketingrészleg vezetőjének; ezenkívül *B. R. Howes*-nak, az adatbázis-előállítás vezetőjének; *Mr. Greenwood*-nak, a korrektúrarészleg vezetőjének; *J. Deaves*-nek, a szerkesztési osztály vezetőjének; *G. Jones*-nak, a szelektív információterjesztés felelősének; *P. Ashton*-nak, az online felhasználókkal való kapcsolattartás akkori felelősének; *J. R. Mahoney*-nak, a rendszerfejlesztés vezetőjének és *J. Pache*-nak, a mágnesszalag-szolgáltatás felelősének, akik valamennyien sokat áldoztak ránk értékes idejükből.

### VÁLAS György: Adatbázis-építés: például az INSPEC

Magyarországon is sürgetően időszzerűvé vált az online kereshető adatbázisok létrehozása és ezzel együtt a meglévő saját referáló folyóirataink szerkesztésének számítógéppel való segítése. Ezért érdemes áttekintenünk az ilyen feladattal régebben foglalkozó szervezetek módszereit. Az INSPEC elismerten a világ legjobb adatbázisai sorába tartozik, ezért előállítási módszereit az elsők között kell tanulmányoznunk. Részletesen áttekintjük az INSPEC adatbázis-előállítási folyamatát és a hozzá kapcsolódó nyomtatott kiadványok számítógéppel segített szerkesztését.

\* \* \*

### VÁLAS, Gy.: To build a database: The INSPEC, for example

It has become a high-priority information project also in Hungary to create online accessible databases and, simultaneously, to aid the edition of existing abstracting journals by computer techniques. Therefore, it is worthwhile to overview the methods developed earlier by organizations on this field. It is well known that INSPEC is one of the best databases so their processing techniques should be in the first place studied thoroughly. The linked processes of database generation and computer-aided edition of printed products of INSPEC are described in detail.

\* \* \*

**ВАЛАШ, Дь.: Создание баз данных: например, INSPEC**

В Венгрии очень назрел вопрос о создании баз данных, доступных в диалоговом режиме, и одновременно с этим об облегчении редактирования собственных реферативных изданий. Поэтому целесообразно сделать обзор методов тех организаций, которые уже раньше занимались такими вопросами. INSPEC относится к лучшим базам данных мира, поэтому, в первую очередь, необходимо изучать методы ее создания. В статье подробно анализируется процесс создания базы данных INSPEC и ход редактирования печатных изданий при помощи ЭВМ.

**VÁLAS, Gy.: Datenbasis-Aufbau, zum Beispiel INSPEC**

Auch Ungarn ist es dringend aktuell geworden, on-line suchbare Datenbasen zustandezubringen, und damit die Verfassung unserer eigenen, vorhandenen Referateblätter mit der Anwendung von Rechenmaschinen zu fördern. Deswegen lohnt es sich, die Methoden solcher Organisationen, die sich mit derartigen Aufgaben längst beschäftigen zu überblicken. Die INSPEC gehört zur Reihe der besten Datenbasen der Welt, so haben wir ihre Herstellungsmethoden unter den ersten zu untersuchen. Der Herstellungsvorgang der INSPEC-Datenbasis und die mit der Rechenmaschine geförderte Verfassung der dazu gehörenden gedruckten Verlagswerke werden ausführlich analysiert.

