

tikus közös nyelvi alapjáról volt szó, aztán az a megvalósíthatatlan gondolat következett, hogy a kettős-pontos osztályozást használják nemzetközi nyelvként. Bár a gépi fordításnak és az információkeresésnek ez a korai kapcsolata nem volt különösképpen gyümölcsöző, meg kell jegyezni, hogy a PRECIS osztályozási rendszer segítségével kapott szövegek már megközelítőbbek a gépi fordítás számára.

A szakértői rendszerek közvetlen alkalmazása az információs munkában az utóbbi idők fejleménye, elsődlegesen az online bibliográfiai adatbázis-keresés területén. A korai próbálkozások határozottan arra mutatnak, hogy a jövő felhasználói sikeres online irodalomkereséseket végezhetnek majd a könyvtárosok és az információs szakemberek segítségével is. Az Egyesült Államokban működik például a PAPERSCHASE — önkiszolgáló lehetőség, nem igényel felhasználói kézikönyvet vagy képzett közvetítőt, ezáltal csökkenti a keresési költségeket. A National Library of Medicine fejleszti a Hepatitisz Tudásbázist, amely hiteles és mérvadó közlések formájában válaszol a kérdésekre (lerövidítve azt a hosszú utat, amit a hivatkozás visszakeresése, a dokumentum megszerzése és elolvasása jelent).

A hagyományos könyvtári munkát szolgáló szakértői rendszerek fejlesztése terén kevés előrehaladás történt, de a szabály-alapú katalogizálási és osztályozási tevékenységek különös figyelmet érdemelnek. A katalogizálás terén két ajánlat merül fel: első lépésként ember-gép interfész, ahol a szellemi erőfeszítés megoszlik az ember és a támogató rendszer között, majd szakértői rendszer teljes körű katalogizálási ismeretekkel, elektronikus kiadói rendszerhez kapcsolva, úgy, hogy amint a szöveg online generálódik, átmehet a katalogizálási eljárás emberi szellemi beavatkozás nélkül. Angliában az Exeter egyetemen egy katalogizálási tudásbázis potenciális szabályforrásaiként az AACR2 és a MARC Manual használatát vizsgálják, kutatják a természetes

nyelvű adatlekérdezés, az intellektuális tudás és az adatbázisból való visszakeresés kérdéseit.

Osztályozási szakértői rendszer is elképzelhető, amely először átrostálná, súlyozná és rangsorolná a leggyakrabban használt és legrelevánsabb kifejezéseket, majd összehasonlítja őket a tudásbázisban nyilvántartott ismerethalmazokkal. Mielőtt az optimális jelzetről dönt a rendszer, az eredményeket az ésszerűség szempontjából még egyszer felül lehetne vizsgálni. Az indexelést és a kivonatkészítést is támogatni lehet a szakértői rendszerekkel. Lehetne tudásbázisokat készíteni, amelyekhez akkor fordulnánk, ha hardver és szoftver kiválasztása és értékelése valamint információs rendszerek tervezése és értékelése terén akarunk tanácsot kérni.

Az ilyen fejlesztéseknek hatása van a könyvtár- és információtudományi képzésre is és az információs szakemberek szerepére is. A szakértői rendszerek használata megszüntethet egyes jelenlegi funkciókat, de új szerepeket is teremthet, vagy legalábbis szabaddá teszi a szakembereket más feladatok ellátására. A vita középpontjában az információszolgáltatás és a tanácsadás közötti különbség áll, az utóbbi több tudást kíván és utat nyit a személyes hajlamoknak. Ezen a ponton azonban alá kell húzni az információs szakember fontosabbik szerepét: ő a kapus az információszolgáltatás területén. Ezt a szerepet elfogadni annyit tesz, mint tudni, hogy az új fogalmak és technológiák asszimilálása vagy elutasítása nemcsak az információszolgáltatások jövőjét, alakítja, hanem meghatározza civilizációnk és kultúránk haladását is.

/ YAGHMAI, N. S. — MAXIN, J. A.: *Expert systems: a tutorial = Journal of the American Society for Information Science*, 35. köt. 5. sz. 1984. p. 297-305./

(Szöllősy Éva)

A CAS adatbázis előállításában alkalmazott technológiák

A CAS adatbázis felépítése

A Chemical Abstracts Service (CAS) a 60-as évek második felében kezdte meg automatizált rendszerének kiépítését. A rendszer inputja bibliográfiai hivatkozásokból, referátumokból és mutatótelekből áll. Arra törekedtek, hogy a dokumentum feldolgozás eredményei a lehető legkorábbi fázisban váljanak géppel kezelhetővé, s — ahol csak lehet-

séges — számítógépes szerkesztési eljárást alkalmazzanak, hogy az emberi beavatkozást az adatok viszonylag szerény körére korlátozhatják. Az egyes kiadványok, indexek vagy szolgáltatások, mint outputok vagy fényszedés révén nyomtatásban, vagy számítógéppel olvasható formában állnak rendelkezésre.

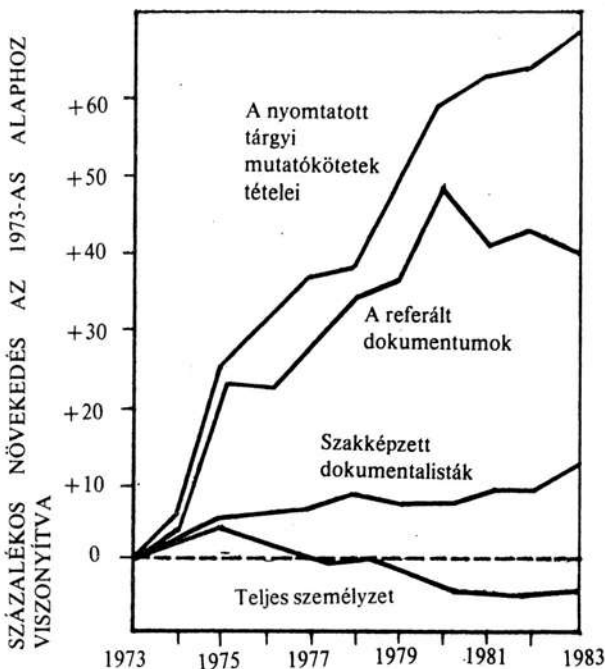
A CAS feldolgozó-szerkesztő rendszerének funkcionális felépítése a következő:

- ◆ a dokumentumok beszerzése
- ◆ bibliográfiai leírás és azonosítás
- ◆ a dokumentumok feltárása (referálás és indexelés)
- ◆ a számítógéppel olvasható adattétel elkészítése
- ◆ a szótár szabványosítása és ellenőrzése
- ◆ szerkesztés
- ◆ a kiadvány anyagának kiválogatása és megszerzése
- ◆ a kiadvány előállítása és postázása.

Jelen cikk az adatbázis felépítésében és karbantartásában alkalmazott technológiával foglalkozik.

A számítógépes technológia hatása a CAS feldolgozására

A számítógép alkalmazásának gazdaságosságát az 1. ábra szemlélteti, a Chemical Abstracts (CA) számára 1973 és 1983 között feldolgozott dokumentumok, illetve a kiadvány tételeinek számának, továbbá a feldolgozást végző munkatársak, illetve a CA teljes személyzete létszámának alakulásával. A növekedés az 1973-as bázis százalékában van kifejezve; volumenét néhány abszolút számmal lehet jellemezni: 1983-ban 451 573 referátumot közölt a CA, s a Chemical Substance Index-ben 2,1 millió, a General Subject Index-ben 1,2 millió indextétel szerepelt, továbbá 900 000 molekula-indextétel, 1,2 millió szerzői és feltalálói név és 75 000 szabadalmi szám. Az online keresés számára még több ponton férhető hozzá az adatbázis.



1. ábra

Míg a feldolgozott dokumentumok száma 41%-kal és a kiadvány indextételei 67%-kal emelkedtek, addig a feldolgozó személyzet létszáma csupán 11,5%-kal. Ha pedig a teljes személyzetet tekintjük, 4,6%-os csökkenés következett be. S mindez úgy, hogy 1973-ban a referátumok 37%-át készítették önkéntes külső közreműködők, 1983-ban azonban már csak 4%-át.

Ugyanakkor a CAS szolgáltatások használói is hasznot húztak a modern technológia alkalmazásából. A szolgáltatások árai a korábnál lassabban emelkedtek, sőt, a dollár vásárlóerejének csökkenéséhez mérten a felhasználó feldolgozott dokumentumként átlagosan kevesebbet fizetett 1983-ban, mint 1973-ban. Növekedett a pontosság és a következetesség, csökkent az átlagos átfutási idő, ami most elérte a 80 napos rekordot, — de a világ több, mint 700 vezető kémiai folyóirata tekintetében ez valamivel kevesebb, mint két hónap. A féléves indextételek előállítása egyharmaddal rövidebb ideig tart ma, mint 1973-ban, s 95 nappal a tárgyidőszak lezárása után már postázhatók.

Online szerkesztés

A CAS feldolgozási menete szerint miután a dokumentalista megismerkedett a dokumentum tartalmával, magnetofonba diktálja a referátumot és a megfelelő indextételeket. Mások írják le géppel olvasható formára a tételeket, amelyeket a központi számítógép dolgoz fel különböző programok szerint. Az online szerkesztés bevezetése előtt papírvonaton korrigálták a szerkesztők a szöveget, amely javítás céljából ismét bejárta a korábban egyszer már megtett utat.

Kísérletképpen 1975-ben először a Chemical Substance Index-nél alkalmazták az online szerkesztést, majd 1977-től lényegében minden szerkesztői munkát így végeznek.

Kezdetben a rendszer három IV-Phase miniszámítógép-terminálcsoportból állt (egyenként 16 képernyős terminál), amelyek egy IBM 370/168 típusú központi számítógéphez csatlakoztak.

A fejlesztés következő szakaszának az volt a célja, hogy tehermentesítsék a központi számítógépet és biztonságosabbá tegyék az üzemeltetést: 48 Beehive B500 programozható képernyős terminált kötöttek két PDP-11/34 miniszámítógéphez, amelyek nagy mennyiségű adatsorokat képesek tárolni. Ez utóbbiak hívták le a központi számítógépből a szükséges adatsorokat, a szerkesztési munkát pedig a terminálokra és a miniszámítógépeken végezték.

A jelenleg folyamatban lévő fejlesztés végén minden dokumentalista irodájában egy DEC PC350 személyi számítógép fog állni. A mintegy 250 sze-

mélyi számítógép a CAS telefonhálózatán keresztül kapcsolódik a terminálközpontokhoz (terminalconcentrator). E központok két PDP-11/44 miniszámítógép és a munkaadások között továbbítják odavissza az információkat a szerkesztési műveletek elvégzése során. A miniszámítógépek a munkába vett adatsorokat egy IBM 370/168AP típusú központi számítógépről hívják le. Az új konfiguráció automatikusan szabályozza e két miniszámítógép terhelését és így nagyobb hatékonyságot eredményez; ugyanakkor a dokumentalisták is több segítséget kapnak a számítógéptől (pl. a tárgyi indexek fájljaihoz közvetlenül férhetnek hozzá).

Az online szerkesztésre való áttérés jelentősen megnövelte a munka hatékonyságát, különösen a referátumok és a tárgyszóindex tételei szerkesztése terén. A három leghatékonyabb szerkesztési programnak a következők bizonyultak: a helyesírási hibákat felderítő program, az automatikus rövidítési program és az általános tárgyi index tárgyszavainak verifikációs programja.

Az online szerkesztés gazdaságosságára vonatkozó feltételezések eleinte csak az irodai segédzsemmélyzet, a papírmunka stb. csökkenését várták. Jelentős többlethasználóként jelentkezett, hogy a dokumentalisták termelékenysége is megnövekedett. A segédzsemmélyzet létszáma 43 főfoglalkozású állás-helynek megfelelő mértékben csökkent. Korábban hetenként 135 000 papír korrektúralevonat készült; ez most 33 000-re csökkent. Az évi megtakarítás 5,5 millió papírlap.

Végül megjegyzendő, hogy a CA referátumainak szerkesztőségi átfutási ideje is négy-öt nappal csökkent.

A szerzői mutató előállítása

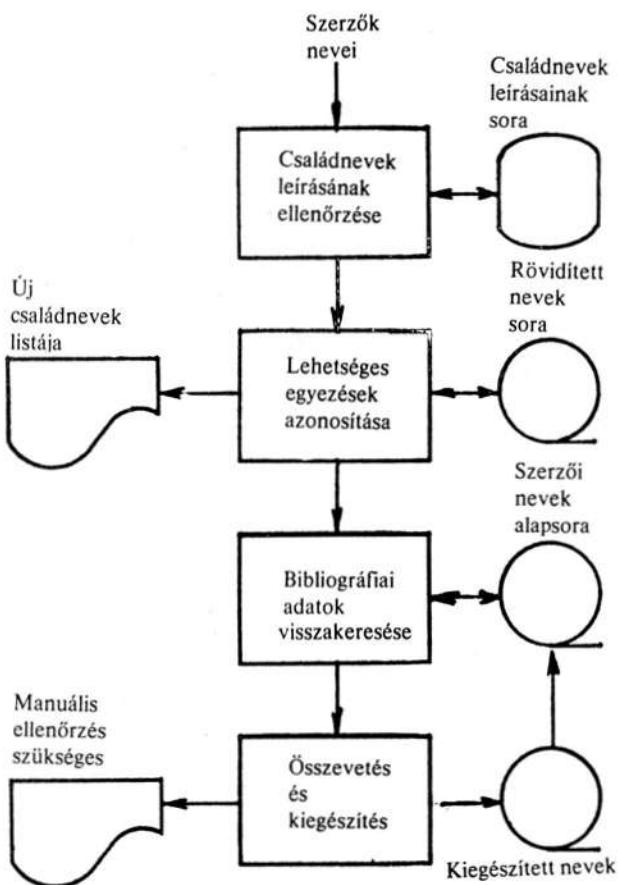
A számítógépes technológiát a CA szerzői mutatójának előállításában is alkalmazták (Author Index Manufacturing System – AIMS), azzal a céllal, hogy a féléves és ötéves mutatók minőségét javítsák, az átfutási időt lerövidítsék és az előállítási költségeket csökkentsék.

A CAS nagy súlyt helyez a szerzői nevek teljes és pontos közlésére annak érdekében, hogy minden cikk és szabadalom szellemi tulajdonosa neve alatt szerepeljen az indexekben.

A CAS a szerzői neveket abban a formában táplálja be az adatbázisba, amelyben a primér irodalomban megjelentek. Ez azt jelenti, hogy előbb-utóbb egy szerző neve többféle alakban (teljes keresztnévvel, csak kezdőbetűkkel stb.) fog szerepelni. A múltban nagy erőfeszítéseket igényelt a nevek kiegészítése és azonosítása.

Az AIMS azt a szellemi folyamatot szimulálja, amelynek során az újabb neveket a szerkesztők a már meglévővel hasonlították össze, s – ha lehetséges – automatikusan kiegészítik őket.

A 2. ábra leegyszerűsítve mutatja be az AIMS működését. A családnevek sora (Surname Spelling File) mintegy 400 000 egyéni szerző és 35 000 testületi szerző nevé tartalmazza a CA utolsó öt évének szerzői mutatóiból. A naponta feldolgozott 4400 nevet a számítógép összeveti ezzel a listával; amelyik névnek nem találja meg pontos megfelelőjét, azt kinyomtatja. Ez az esetek 6%-ában következik be.



2. ábra

A szerzői alapsor (Author Master File) több, mint 2,2 millió, az elmúlt öt évben referált dokumentum bibliográfiai adatait tartalmazza; ebben 5,5 millió szerző szerepel. Ez a sor lehetővé teszi a beérkező szerzői nevek pontos azonosítását és kiegészítését. Ha a beérkező új név nem szerepel a sorban, ellenőrzés és azonosítás végett a gép kiírja, s mindazokat a neveket is, amelyek vele összefügghetnek. Ez a folyamat a szerzői alapsor kiegészítését is szolgálja.

A rövidített nevek sora (Abbreviated Name File) a szerzői alapsorban szereplő szerzők családnevét s

két első keresztnévük kezdőbetűjét tartalmazza, s a névkiegészítés első lépcsőjeként használják.

Az AIMS bevezetése a szerzői nevek 91%-a esetében feleslegessé tette a manuális szerkesztést, s mintegy két hónappal lerövidítette egy szerzői mutatókötet előállítását; itt az emberi munka 27%-kal csökkent. Különösen az ötéves szerzői mutató esetében látványos a megtakarítás: az 1977–1981-es kötethez az előzőkhöz szükséges manuális munkának csupán 4%-a volt szükséges, noha a feldolgozott dokumentumok száma 24%-kal több volt. A manuális szerkesztés időtartama két évről három hónapra csökkent, s megbízhatóbb és pontosabb adatokat eredményezett.

Az online szerkesztés és a szerzői mutató előállítása mellett a CAS a számítógépet több más, hasonló területen is alkalmazza a munkatermelékenység fokozására, az átfutási idő lerövidítésére és a minőség javítására.

A számítógépes technológia további alkalmazása

1984-ben a CAS további lépéseket tesz avégett, hogy még inkább hasznosítsa a számítógépes technológiát adatbázisa felépítésében.

Az 1960-as években hozták létre a CAS Registry System-et: 1965 januárja óta a CA-ban indexelt vegyületeket ebben tartják nyilván. Most a CAS egy több évre szóló programba kezd azzal a céllal, hogy az 1965 előtt indexelt vegyületeket is beiktassa ebbe a regiszterbe. Első lépésként a nyomtatott Collective Formula Index adatait géppel olvasható formára hozzák egy optikai leolvasó berendezés (Kurzweil Data Entry Machine) segítségével. Ezután következhet a vegyületek automatikus összevetése a CAS Registry File-ban szereplőkkel. Ha nincs egyezés, a vegyi szerkezetet automatikusan beillesztik az adatbázisba. Csak azok a tételek kerülnek a kémikusokhoz azonosításra, amelyeket a számítógép képtelen volt feldolgozni.

Ez a fejlesztés is jól mutatja a CAS-nak azt a törekvését, hogy az emberi tevékenységet megszabadítsa a rutinszerűen ismétlődő munkáktól és helyette a problémák megoldására hasznosítsa.

/ WEISGERBER, D. W.: Applications of technology to CAS data-base production = Information Services Use, 4. köt. 5. sz. 1984. p. 317–325./

(Papp István)

Hogyan tart lépést a Chemical Abstracts Service a japán vegyipari technológia fejlődésével?

Az Egyesült Államokban az utóbbi időben kormány szinten foglalkoznak azzal a problémával, hogy az amerikai ipar nem tudja kellőképpen kiaknázni a japán technológia és kutatás növekvő eredményeit. Ennek az lehet az oka, hogy a japán tudományos és technikai információkhoz nem lehet könnyen, gyorsan hozzájutni az USA-ban. Jelen közlemény azt vizsgálja, hogy az egyik legjelentősebb USA-adatbázis, a Chemical Abstracts Service (CAS) hogyan oldja meg a japán kémia és vegyipari technológia terén elért fejlődés eredményeinek feltárását.

A CAS a világ 133 országából származó 12 ezer tudományos és műszaki tárgyú időszakos kiadvány, 26 országban kiadott szabadalmak, a világ minden tájáról érkező konferenciakiadványok, disszertációk, kutatásjelentések és könyvek feldolgozását végzi. 1983-ban a CAS munkatársai 451 753 angol nyelvű kivonatot készítettek, ezenkívül 95 811 ekvivalens szabadalmat azonosítottak a Patent Indexen keresztül.

A japán vegyipari technológia fejlődése

Japánnak a kémiai és vegyipari technológiai kutatások területén betöltött rendkívül jelentős szerepét bizonyítja a CAS által 1983-ban feldolgozott kémiai folyóirat-irodalom forrásainak elemzése. A CAS 1907 óta kivonatolja és indexeli a japán kémiai és vegyipari technológiai szakirodalmat. A feldolgozandó kiadványokat — előfizetés csak csere révén — a CAS közvetlenül azoktól a japán akadémiai, kereskedelmi és kormányzervektől szerzi be, amelyek ezeket megjelentetik. A többnyire még meg sem jelent kiadványok légipostán érkeznek a CAS-hoz (pl. tördelt levonat formában).

Az 1. táblázat a japán kémiai és vegyipari technológiai szakirodalom számbeli növekedését mutatja 1907-től 1983-ig.

Az elmúlt 3 évtized során a japán tudományos közlemények (folyóiratcikkek, konferenciaanyagok és kutatási jelentések) száma gyorsan és egyenletesen növekedett. A 2. táblázat azt mutatja, hogy