

AZ AUTOMATIZÁLT TUDOMÁNYOS DOKUMENTÁCIÓ ÚJ IRÁNYZATAI

Georges SAG

Nemrég megjelent cikkünkben (1.) ismertettük szovjet kollégáinknak a dokumentáció szervezését illető kívánságait és összehasonlítottuk Bouruy-Smar'jan strukturális terveit a hasonló nyugat-európai tervekkel. Megállapíthatjuk, hogy a két rendszer közös nevezője a szakemberek hiánya, olyan szakembereké, akik "meg tudják érteni a használók problémáit, s el tudják vezetni őket a számukra szükséges információkhoz, hála az információkat tartalmazó dokumentumok ismeretének és a megfelelő információtároló rendszernek, amelynek segítségével bármikor azonosítani lehet azoknak a dokumentumoknak az adatait, amelyek válaszolnak a használók által feltett kérdésekre." (2).

A dokumentáció automatizálása, amelynek eszméje 30 éve vetődött fel Franciaországban (3), ma már - legalább is részben - meg tudja oldani a problémákat. Először is a dokumentumok referenz-adatainak azonosítása mechanizált és még inkább elektronikus rendszerek segítségével, a hagyományos kézi módszerekhez képest jelentős időmegtakarítással végezhető el. A dokumentumok tartalmának elektronikus számítógépek segítségével történő tárolása pedig most van a megvalósulás útján.

Alábbiakban az automatizált műszaki-tudományos dokumentációban jelentkező két irányzatot ismertetünk röviden:

A. Az emberi agy által végzett elemző munka eredményeinek magneses adathordozókon való tárolása és megfelelő memóriákban való visszakeresése.

B. Teljesen automatizált adattárolás.

Különösen nagy súlyt fektetünk e két fejlődőben levő módszer gazdasági kérdéseire.

⁺ A cikk a "Chimie et Industrie" c. folyóirat 1968. 1. számában (94-100. p.) jelent meg, itt mondunk köszönetet a szerzőnek, aki megküldte a Könyvtártudományi Szakkönyvtárnak az eredeti szöveget.

Dokumentációs visszakeresésnek valamely dokumentumgyűjteményben levő s a használó által feltett kérdésre válaszol. Dokumentumok kiválogatását értjük. Ez a válogatás vonatkozhat a már meglevő állományra, ez esetben retrospektív visszakeresésről beszélünk, vagy pedig ennek az állománynak periodikusan növekvő anyagára, ez esetben általában információszésugárzásról van szó, a speciális igényű használók szükségleteinek ("az érdeklődés profiljának") megfelelően.

A.

Az automatizált dokumentációs visszakeresés céljára az anyagot tárolni kell. A dokumentációs anyagot háromféleképpen lehet tárolni:

1 A ma még mindig legelterjedtebb módszer az egyes dokumentumok kézi indexelése, magukból a dokumentumokból vett kulcsszavak vagy deskriptorok segítségével, amelyeket esetleg egy már összeállított és szabványosított deskriptor-jegyzékkel való egybevetés útján folyamatosan kiegészítünk. Ezt a megoldást fogadtuk el az Oréal dokumentációs szolgálatában. (4). E módszer különösen jól beválik a mi szakterületünkön, a kémiában. A helyzet az, hogy bár van nemzetközi nomenklatura, a kereskedelmi vagy szakmabeli szavak használata a kiadványokban igen elterjedt, márpedig ha a szabványosított deskriptor-jegyzékben vagy szakszótárban a hidroxibenzol szó szerepel, a számítógép nem tudja azonosítani a fenol vagy a szénsav szóval, sem a tárdás, sem a visszakeresés folyamatában, hacsak e három szót fel nem veszik szinonimaként a szókészletbe.

A deskriptorok közötti logikai kapcsolatokat is kódolni lehet a tárolás során (és fel lehet használni a keresés folyamatában). Erre a célra két többé-kevésbé kidolgozott rendszert javasolnak:

a) A "szerep-indikátorok" a deskriptorokhoz kapcsolt olyan szimbólumok, amelyek segítségével meg lehet különböztetni például a nyersanyagokat, az előállított termékeket, melléktermékeket, selejtet, oldószereket, katalizátorokat, a tanulmányozott változókat, alkalmazott eljárásokat stb.

b) A retesz ("link") az egyazon dokumentum deskriptorai közötti hiányzó interpunkciót pótolja. Így például egy olyan tanulmányt, amely forrástárral, hegesztéssel vagy szegecseléssel egyesített különböző fémeknek csavarással, nyújtással vagy összenyomással szembeni ellenállásáról szól, indexelhetünk a tanulmányozott fémeket jelölő kulcsszavakkal, ugyanannak a "link"-nek a hozzáadásával, míg a tanulmányozott változókat jelölő deskriptorokat más, azonos "link"-kel látjuk el.

c) A Syntol nyelv (5) kódolási egységként szintagmákat használ, vagyis a kapcsolatot kifejező indikátorokkal összefűzött terminusokat. Így például a Ri (a, b,) szintagma azt jelenti, hogy az a szó Ri kapcsolatban van a b-vel, de b nincs okvetlenül Ri kapcsolatban a-val.

d) Mi a mi rendszerünkben a "reteszelés" alkalmazzuk, de - az American Institute of Chemical Engineers-sel ellentétben - a dokumentumok tárolásának, nem pedig a deskriptoroknak a színjén: vagyis a memóriába tárolást egy durvább osztályozás előzi meg.

A deskriptorok közötti logikai kapcsolatokat a kapcsolatban levő s egyazon dokumentum tartalmát kifejező kulcsszavak közötti helykihagyással kódoljuk. Ezt az eljárásmodot "mellékrendelő szintaxis"-nak nevezzük.

E logikai és/vagy szintaktikai kapcsolatok bevezetésének módszerei azt célozzák, hogy a memória megkérdezésekor korlátok közé szorítsuk a "zajtényező"-t, vagyis kiküszöböljük a számítógépes válogatásból azokat a dokumentumokat, amelyeket ugyan indexeltünk a feltett kérdés deskriptoraival, de amelyek nem felelnek a feltett kérdésre. A "zajtényező" tipikus példajaként említhetjük az alábbi kódolt dokumentum kiválasztását.

Gyártás
Étiléter
Éterré alakítás
Etilalkohol
Katalizátor
Kénsav

a kénsav gyártásánál használt katalizátorokra vonatkozó keresésnél, a következő kérdéssel: Gyártás - Kénsav - Katalizátor.

A zajtényezővel ellentétben "csendnek" nevezzük, ha a memóriához intézett kérdésre választ adó dokumentumot a számítógép nem választja ki. A "csendnek" az az oka, hogy a kódolás és a kérdés nem ekvivalens deskriptorokkal történt. F. Levery szerint (6) különösen akkor fordul ez elő gyakran, ha a használó nem tudja kérdését szabatosan meghatározni és keresése tárgyát szakszerű terminusokban kifejezni.

G. Salton, a Medical Literature Analysis and Retrieval System ("MED-LARS") és a National Aeronautics and Space Administration ("NASA") módszereit kommentálva (7) - ezek kézi indexelésű rendszerek - ugyancsak erre a következtetésre jut: "Ha a használó nem ismeri eléggé a kódolás elméletét és gyakorlatát, nem tudja, milyen kulcsszavakat használjon és ezeket hogyan kombinálja". Végül hozzáteszi: "ha a dokumentáló nem ismeri a használók prob-

lémáit és az általuk feltehető kérdések típusait, nem tud a későbbi retrospektív keresés céljának megfelelő deskriptorokat választani".

A "csend" elkerülésének legegyszerűbb módja, hogy több, különböző paraméterekkel megfogalmazott kérdést teszünk fel. Saltun egyébként azt mondja (7), hogy tökéletes választ nem remélhetünk egyetlen kérdés feltevésével.

Más, jobban kidolgozott rendszerek ugyancsak arról ismeretesek, hogy a kérdezés során igyekeznek a "csendet" limitálni. A MEDLARS és a NASA rendszerekben, valamint a Syntol nyelvben hierarchikus kapcsolatba hozzák a deskriptorokat, paradigmaticus fogalmi fákat alakították ki, a specifikus fogalmakról utalással az általánosabb kulcsszavakra és vice versa (hogy így korlátozzák a válaszok tuburjánzását). Tegyük fel, hogy a memóriában a hollók irtására alkalmas mérget keresik, a Méreg, Pusztítás és Holló kérdésfeltevésével. Ha nincsen olyan dokumentum, amely e deskriptorokat magában foglalná, az elektronikus számítógép változtat a kérdésen, a kódolt hierarchikus kapcsolatok segítségével a "Holló" fogalmát "Madár"-ra változtatja és így növeli a kielégítő válasz lehetőségének kilátásait.

2 Összefoglalásoknak (rezüméknek) - amelyeket, akárcsak a kulcsszó-jegyzékekét, szakemberek készítettek - gépi tárolása szintén meglehetősen elterjedt módszer. Ilyen például a Gaude féle I. B. M. módszer.

3 Végül rezümék és deskriptorok egyidejű tárolása a harmadik módszer dokumentumoknak a számítógépbe való kézi bevitelére. Ezt a módszert alkalmazza egy idő óta többek között a Chemical Abstracts Service ("C. A. S.").

Az így felépített mágneses memóriában való keresés vitathatatlanul több problémát vet fel, mint a kizárólag kulcsszavakkal kódoló rendszerek; kiterjedt szinonimák, paradigmaticus fogalmi fákat magukban foglaló szótárak, a szemantikai rokonság szerint csoportosított szójegyzékek (például a kémiában homológ sorozatok) használata itt kötelező.

"Súlyosabb ismérvek" (olyan kifejezéseknek tulajdonított értékőbbslet, amelyek előfordulhatnak a rezümékben) növelhetik a rendszerek használhatóságának esélyeit.

Rezümékben található szótöredékek keresése - a C. A. S. módszer által fel felkinált egyik lehetőség. Ennek segítségével automatikusan összeállíthatók olyan csoportok és kiválogathatók azok a cikkek, amelyekben benne van például a "Steron" szótöredék. Ilyenek pl. Aldosteron, Andosteron, Corticosteron, Nortestosteron, Testosteron stb.

A fenti három módszer által felvetett gazdaságossági problémák legsúlyosabbja nyelvi jellegű. Bár az angol nyelv jelenleg tulsúlyban van a világ kémiai szakirodalmában (a kiadványoknak kb. 60 százaléka angol nyelvű), jelentős mennyiséget képviselnek például a japán, orosz, német nyelvű kiadványok, sőt rövidesen a kínai nyelvűek is. Ha egy dokumentációs szolgálat valamennyire is teljességre törekszik, nem hanyagolhatja el az ezeken a nyelveken megjelent publikációkat.

Felesleges itt hangsúlyozni, milyen nehéz nálunk olyan tudományos dokumentálókat szerezni, akik kellő szinten tudják a kémiai szakirodalom feldolgozásához szükséges (2) legnélkülözhetetlenebb nyelvet: az angolt és a németet.

Ezzel a nyelvi igényrel szemben lehetne itt azzal érvelni, hogy olyan ki-tűnő referáló lapok vannak a kémiában, mint a C. A. B., a Chemisches Zentralblatt vagy a Referativnij Zsurnal. Viszont ezek a kiadványok a kémia szakterületének csak általános szempontok szerint kiválogatott irodalmát fogják át. Így tehát - P. R. Brygoo-val (8) - meg vagyunk győződve, hogy az enciklopédikus jellegű dokumentációs szervezetek nem mindig kielégítőek a messzemenően specializált dokumentációs központok vagy szolgáltatások szempontjából. Másrészt azt is hangsúlyozni kell, hogy a rezümé sohasem pótolja az eredetit.

Azonkívül ezekben a referáló lapokban az eredeti publikációról gyakran nagy késéssel jelennek meg a kivonatok, ami súlyos szavatossági hiba, különösen a szabadalmakkal foglalkozó dokumentációs szolgálatok szempontjából.

Mindezek következtében úgy gondoljuk, hogy az emberi értelem által végzett elemzés segítségével történő gépi tárolást az eredeti dokumentum elolvasása és megértése kell hogy megelőzze. Saját 20 éves tapasztalatunk alapján úgy látjuk, hogy a kémiai dokumentációban egy gyakorlott szakember 8 órás munkanapon átlag 10 eredeti dokumentumot elemezhet, számításba véve a dokumentum hosszúságát, nehézségét és új fogalmakban való gazdagságát is (új deskriptorok bevezetését a szójegyzekekbe). Ebből az következik, hogy az elemzést és kódolást végző dokumentáló átlag 2500 dokumentumot dolgozhat fel évenként. Ha egy dokumentáló munkabére mondjuk kb. 48 000 Ft (a rezsit is beleértve), akkor egy olyan dokumentum tárolása, amelyhez nincs szükség külön fordítóra, kb. 20 Ft-ba kerül.

A memóriában való tárolás, átlagban 12 deskriptornyi mélységű elemzést és 3 sorból álló címléírást és hivatkozásokat véve alapul (átlagosan 20 n-ből álló sorokkal), a lyukasztás és az azonosítás költségei (a gépek bérösszegét is beleértve) kb. 1.70 Ft-ra rúgnak. A számítógépbe való tulajdonképpeni betáplálás, 400 dokumentumból álló adagokban, I. B. M. 360/30-as, viszonylag egyszerű programozású gép alkalmazásával, a szalagok vagy lemezek periodikus módosítását - a megelőzőleg tárolt dokumentumokkal kapcsolatban elköve-

tett tévedések, hibák kijavítását, hiányok pótlását - valamint a különböző szövegjegyzékek kinyomatását is tekintetbe véve - valamivel több mint 1.00 Ft költséget jelent. Egy-egy dokumentum tárolásának költsége tehát - félautomata módszerekkel - globálisan kb. 24 Ft-ra tehető.

Egy másik gazdaságossági tényező, az adattárolás elsőként javasolt megoldása érdekében - rezümék helyett deskriptorok használata, esetleg kulcsszavakkal - a lyukasztási munkák jelentős részének kiküszöbölése, hacsak nem (Flexowriter típusú) lyukszalagos írógépet használunk, ez esetben azonban tekintetbe kell venni a gépek amortizációját és a lyukasztókat és azonosítókat helyettesítő elektronikus eszközök bérköltségeit. Ezek az eszközök viszik át a lyukszalag tartalmát ("szalag - kártya konverter") az információfeldolgozó számítógép központi egységébe.

B.

Ami az ismeretek teljesen automatizált tárolását illeti, három teljesen különböző rendszert kell megkülönböztetni. Ezek:

1 A permutált indexek.

Ezt a rendszert, amelyet főleg "key word in context" (KWIC) vagy "key word in title" (KWIT) - szöveggörnyezetes index - néven ismerünk, a mi szakterületünkön a C. A. S. alkalmazza néhány év óta.

Permutált index készítésénél általában a cikk vagy a szabadalom címéből kiemelt egész szövegrészeket visznek be a számítógépbe, amely e szövegrészeket összehasonlítja egy "ellen-szótárval" vagyis a permutációból kizárandó szavak jegyzékével, a fennmaradó értékes (releváns) szavakat betűrendezi, s a szövegrészben szereplő szomszédos szavakkal együtt nyomtatott tárgymutatót készít belőlük, amelyben minden értékes kifejezést a szövegrész többi szava magyaráz. Ebben a rendszerben a kétértelmű vagy túl terjedelmes címek esetében a dokumentálónak kézi erővel kell elvégeznie a szükséges kiegészítést, illetve pontos megfogalmazást. Ne felejtjük el, hogy a mi iparágunkban a dokumentálókhoz intézett kérdések többnyire valamilyen vegyület vagy vegyület-csoport készítésére vagy specifikus felhasználására vonatkoznak, és hogy a feltett kérdésre választ adó dokumentumok címe gyakran nem nevezi meg a keresett vegyületet.

Véleményünk szerint az elvben könnyen elkészíthető permutált indexek valóban hatékonyan csak akkor használhatók, ha szakemberek szellemi munkája fekszik bennük, ami viszont majdnem olyan költséges megoldás, mint a deskriptor-jegyzékek készítése.

2 Automatikus kivonatkészítés statisztikai módszerekkel.

Az ismeretanyag mágneses memóriában való tárolásának problémáit radikálisan csak magának a dokumentumnak teljes egészében a gépben való tárolása oldja meg.

A kivonatkészítést az eredeti dokumentumok alapján, statisztikai módszerekkel, elektronikus számítógép segítségével, első ízben H. P. Luhn javasolta. (9). Ezzel a módszerrel a gép mágneses adathordozóra írja át az elemzendő (referálandó) dokumentumot, majd kiszűri egy előre elkészített szójegyzék segítségével az értéktelen szavakat (névelők, kötőszavak stb.), a megmaradtakat betűrendezi, az előttük vagy utánuk álló interpunkció feltüntetésével, s a szalagon feljegyzi, hogy az egyes szavak hol szerepelnek a szövegben. Ezután csökken-ti a szinonimák számát, összehasonlítva bizonyos meghatározott számú szókezdő betűt. Azonosság esetén az illető szavakat szinonimáknak tekintik (+ 5 százalékos pontosság). Ha az értékes szavak nem fordulnak elő bizonyos előre meghatározott gyakorisággal, ugyancsak kiszűri őket. Egy példaként említett 2326 szóból álló cikk 39 különböző, rendkívül értékes szóból áll. A gép ezután ki-keresi azokat a mondatokat, amelyek a legnagyobb számban, a legsűrűbben tartalmazták a különösen értékes szavakat. A rezümé úgy készül, hogy a gép kinyomatja azokat a teljes mondatokat, amelyeknek érték-faktora meghaladja (tul-lépi) az előre felállított érték-küszöböt.

A módszert irányítani is lehet: az értékes szavak jegyzékét össze lehet hasonlítani a használtól különösen érdeklő szavak jegyzékével (súlypont jelzése), s ezzel növelni lehet az ilyen szavakat tartalmazó mondatok "érték-faktorát".

E rendszernek francia nyelvű dokumentumok esetében való alkalmazása - még a különösen értékes szavak statisztikai becslését megelőzően - e szavak közötti szinonima (rokonértelmiség) felkutatását teszi szükségessé, mivel a francia nyelv stilisztikai követelményei nem engedik meg az azonos kifejezések túl gyakori ismétlését.

Luhn módszerét, tudásunk szerint, eddig nem használták fel a kémiai do-kumentációban. Attól félünk, hogy a Luhn által idézett cikkeknél lényegesen hosz-szabb tanulmányok megoldhatatlan problémákat vetnének fel.

3 Automatizált kivonatkészítés nyelvi elemzéssel.

Egy másik automatizált kivonatoló rendszer a SMART (Salton's Magical Automatic Retriever of Texts), Salton és munkatársai rendszere (10). E mód-szerrel, mint Luhn módszerével, a szöveget az elektronikus gép teljes egészé-ben belyukasztja, majd elemzi. Az elemzés eredményeképpen több különböző mágneses szalag készülhet, közülük azonban csak az első kötelező:

a) Az első szalag úgy készül, hogy a gép a szöveg minden szavát betűről betűre összehasonlítja egy ugyancsak előre elkészített szuffixum-szótárral. Az azonosított szótöveket és szuffixumokat számkódokkal jelöli s ezekhez hozzáfűzi az ugyancsak előre jegyzékbe foglalt s zintaktikai és/vagy szemantikai kódokat.

A SMART, jelenlegi kísérleti stádiumában, a dokumentológiában, a számítógép-elméletben és az aerodinamikában használható fel: ezeknek szójegyzékei mintegy 1000 angol szótövet tartalmaznak, továbbá 150 szuffixumot és 200 szemantikai (jelentéstani) jelzést, valamennyit kódolva, szakterületenként 1000-en aluli számú dokumentumot véve alapul.

A nem azonosított szótöveket a számítógép mint ismeretleneket kinyomatta, kézi feldolgozás céljaira, vagy pedig mint egy "vak szótár" szavait dolgozza fel.

Magától értetődően az ilyen automatikus elemzés útján nyert kódszámok a szinonimáknak tekintett kifejezések esetében azonosak.

b) A másik szalag úgy készül, hogy a fenti elemzés eredményét a gép összehasonlítja a kódolt fogalmak hierarchikus kapcsolatairól készült, előre összeállított jegyzékkel. Nagyon hasonlít ez az automatizált művelet a kézi dokumentációs módszerek utalócéduláinak használatára. Így például a Hollók-ról ilyen utalókat találhatunk: Id, még; Varjak, Verebek, Madarak; vagy a Halogén elemeknél ilyeneket: Id, még; Fluor, Klór, Bróm, Jód. Az első példa esetében a hierarchikus kiterjesztés az átfogóbb fogalom irányába történt, a második esetben az alárendelt fogalmak irányában; lehetséges egy harmadik eset is: a mellérendelt fogalmak, a paradigmatis fán laterális irányba történő kiterjesztés, mint például a következő esetben: Kivonat, Id, még; Összefoglalás, Referátum, Tömörítvény. A laterális kiterjesztés egyik lehetséges módja a hagyományos dokumentációs módszernél a lásd-utalók használatának felel meg.

c) A harmadik szalag úgy készül, hogy a gép minden szótőről megállapítja, milyen gyakran fordul elő a kielemezett mondatokban és az egész dokumentumban. Ilyen "szó - mondat" egybeesési táblázatok készülhetnek szó-párokról vagy a még fontosabb szócsoportokról is. A dokumentológiából kielemezett gyakran előforduló szócsoporthoz egyik példája: Data processing, Processing of files, Tape handling procedure. Magától értetődik, hogy a Procedure és Processing szavaknak ugyanaz a szótó-kódja. Az ilyen kifejezéseket a szalagon ugyanazok a számjelek képviselik. Ugyanezen a szalagon a számítógép korrelációs koefficienseket (kapcsolati együtthatókat) állíthat elő a szavak között, vagy szó-párok, szókapcsolatok és egyszerű kifejezések közt. Végül a rendszer összeállíthatja a mondatok "értékességi együtthatóit" is, kikeresve azokat a mondato-

kat, amelyek az értékes kifejezéseket statisztikailag a legnagyobb számban tartalmazzák, csatlakozva ily módon Lunn módszeréhez.

d) A negyedik szalag úgy készül, hogy a gép szintaktikai elemzést végez az a) pont alatt leírt szalagon (a szótövek kódjai és a szintaktikai kapcsolatok feltüntető szuffixum-kódok táblázatán), vagy pedig a c) alatt leírt szalagon, amelyen a legmagasabb érték-egyűtthetőket tartalmazó mondatok vannak. A szintaktikai elemzés abból áll, hogy a gép összehasonlítja a mondatok szavait a lehetséges szintaktikai szerkezetek jegyzékével. Ez az összehasonlítás igen sok elméletileg elfogadható szintaktikai szerkezetet eredményezhet, amelyek között a számítógép nem tud választani. Ezenkívül ez az elemzés gépi viszonylatban igen hosszú időráfordítást igényel, ezért a rendszer hívei jelenleg nem gondolnak a szintaktikai elemzés felhasználására (11).

e) Végül az 5. szalag úgy készül, hogy a gép statisztikailag vagy hierarchikusan elemzi az egyedi fogalmak (a dokumentumokból nyert szópárok és szókapcsolatok) közötti kölcsönös összefüggést, az egész dokumentumgyűjteményre vonatkozóan. Az elemzés eredményeként előálló fogalmi csoportokat fogalmi vektoroknak (fogalomhordozóknak) hívják. A számítógépből automatikusan nyert ilyen fogalmi vektorokat úgy tekinthetjük, mint a "reteszelés"-t a dokumentumoknak a szintjén, amelyet rendszerünkben manuálisan kell elvégezni. (4).

A félig automatizált dokumentációs visszakeresésben - deskriptorok használatával - a számítógépnek a kérdések valamennyi kulcsszavát össze kell vetnie valamennyi deskriptorral, amely az összes tárolt és az invertált vagy direkt módon hozzáférhető kartotékban nyilvántartott dokumentumban szerepel. Ha nagy a dokumentumgyűjtemény és mély az elemzés, ez az összehasonlítás túl sok időt vesz igénybe, különösen direkt módon hozzáférhető kartoték esetében. (12). A SMART-rendszerben ki lehet válogatni az adott kérdésre válaszoló dokumentumokat a dokumentumgyűjteményből nyert fogalomhordozóknak a kérdés fogalomhordozójával való összehasonlítása után.

Valamennyi szótár, szójegyzék, amely a különböző tároló szalagok készítésénél szerepel, a SMART-rendszerben automatikusan állítható elő, az illető tudományágat reprezentáló dokumentumanyag alapján. Meg kell azonban jegyezni, hogy a kísérletezés alatt álló 3 tudományág tezaszuszait jelenleg még kézi módszerekkel gyűjtik, ahogy erről 1967 nyarán Salton doktornál tett látogatásunk alkalmából meggyőződhetünk.

A memóriához intézendő kérdések elemzése természetes nyelven készült szövegek alapján történik, ugyanolyan mágneses szalagok készítésével, amelyeket a tárolásnál ismertettünk.

Ami a teljesen automatizáltnak nevezett módszerek gazdaságosság szempontjából való értékelését illeti, el kell ismernünk, hogy e rendszereknél a szakemberek közreműködése a minimálisra csökken. Csak a válaszok értékelése marad manuális munka. A SMART-rendszer egyébként a használó-számítógép párbeszédet javasolja, hogy így - a kérdések folyamatos módosításával - fokozatosan tökéletesítse a válaszokat.

Teljes dokumentumszövegeknek a memóriában tárolása ellenben komoly nyelvi nehézségeket támaszt. Ha minden dokumentumot saját eredeti nyelvén tárolnánk, a kérdéseket is annyi nyelven kellene felvetni, ahány nyelv a memóriában szerepel. Az egyes nyelvek szótöveit, szuffixumait, prefixumait (pl. a német nyelvben), szintaktikai kapcsolatait előzetesen kódolni kellene. A mi nyelvünkétől eltérő írást használó nyelvek dokumentumait lyukasztás előtt transliterálni kellene. A bonyolult kémiai képleteket, ábrákat és grafikonokat ugyancsak át kellene írni lineáris írásra. Végül, ha egyetlen nyelvet akarnánk használni, a képesített szakfordítók egész hadát kellene alkalmazni még egy szerényebb dokumentációs szolgálatnál is.

Az olyan nagy volumenű programok, amilyen a SMART (150 000 utasítás egyedül angol nyelven) hatalmas számítógépek használatát teszik szükségessé (amilyen például az I. B. M. 7094, I. B. M. 360-40 vagy ezekkel egyenértékű gépek), hosszú gépidővel, ami a dokumentumok feldolgozásának önköltségét lényegesen emeli. Teljes dokumentumok lyukasztási-azonosítási költségei (fordítási és/vagy transliterálási költségek nélkül), 4 lapnyi átlagos hosszúságot véve alapul (bibliográfiai adatok nélkül), laponként 70 sort számítva, 50 jelből álló 2-2 hasámban, körülbelül tizszeresét jelentenék a deskriptorokkal történő félig automatizált tárolás költségeinek.

— 0 —

Összefoglalva, úgy véljük, hogy a munkapiac jelenlegi helyzetét, valamint az elektronikus számítógépek bérköltségét tekintetbe véve, a teljesen automatizált dokumentációs rendszerek ma még nem versenytársai azoknak a módszereknek, amelyek a tudományág szakembereinek közreműködését teszik szükségessé a tárolásnál és a kérdések kiválogatásánál. Ha ellenben a félautomata módszereket a hagyományos dokumentációs módszerekkel hasonlítjuk össze, úgy látjuk, hogy a lyukasztás és a számítógépben való tárolás többlet-költségeit bő- ségesen kompenzálja a szakemberek idejének megtakarítása a retrospektív vizs- zakeresésnél és/vagy az érdeklődés profilja szerinti információ-szétsugárzásnál.

Fordította: Molnár Ágnes

Bibliográfiai hivatkozások

1. SAG, G.: La documentation technique en Union Soviétique. Chimie et Industrie, 1967. 97. (no 3), 429-430.
2. FOUQUET, M. L.: Éléments pour une définition de la profession de documentaliste et de son profil en économie. Association des Documentalistes et Bibliothécaires Spécialisés (A. D. B. S.), Groupes d'Études "Économie Générale" 21 pp. Sept. 1966.
3. MAYOR, Y.: Étude des possibilités d'application des machines à statistiques aux recherches bibliographiques. Chimie et Industrie, 1936, 35, (no. 2), 458-466.
4. CHARLE, R. - SAG, G.: Essai d'automatisation de la documentation cosmétologique. Documentaliste, A. D. B. S. Informations, 1966. No. Spécial, 107-114.
5. CROS, R. C. - GARDIN, J. C. - LEVY F.: L'automatisation des recherches documentaires. Ed. Gauthier-Villars, Paris, 1964, 260 pp.
6. LEVERY, F.: La documentation automatique. Concours médical, 1967. 89 (no 1), 97-109.
7. SALTON, G.: Diffusion de l'information et les systèmes automatiques de documentation. Inst. Electr. Electron. Engineers Proc., 1966, 54, (no 12), 1663-1678.
8. BRYGOO, P. R.: Veut-on tuer les Centres Spécialisés? Expansion Rech. Sci. 1963 (18), 21-23.
9. LUHN, H. P.: Création automatique d'extraits documentaires, I. B. M. Jour. Res. Devel., 1958, 2. 159-165.
10. SALTON, G.: Mémorisation et sélection documentaire, Rapport No ISR - 7 à la National Science Foundation, Harvard Computation Lab., 1964.
11. SALTON, G.: Mémorisation et Sélection documentaire, Rapport No ISR - 11 à la National Science Foundation, Cornell Univ., 1966, et communication personnelle.
12. BRETON, GUEDENEY, RAGU, SAG: Association Française d'Informatique et de Recherche Opérationnelle, Groupe d'Études "Mémoires de Masse". Commission IV, Recherche de Documentation. Séance du 15.2 1967.