

Kollacionálás elvén alapuló gépi adatvisszanyerés

OROSZ GÁBOR

Az irodalomfeltáró- és visszakereső eljárásoknak jól elhatárolható csoportját alkotják azok, amelyeknél az adattároló egységek (nevezzük ezeket lapoknak, bár nem mindig azok) a dokumentum jellemzőihez vannak hozzárendelve. A lapok megnevezését, felzetét a dokumentumot karakterizáló fogalmak, descriptorok adják, míg az egyedi adatok felvételére szolgáló rovatokba vagy jelmezőbe a felzetben szereplő fogalommal jellemezhető dokumentumok azonossági számai kerülnek. Az információszolgáltatás céljából végzett visszakeresés a megválaszolendő téma jellemzőinek megfelelő lapok felhasználásával történik. Ezeknek rovataiban, illetve jelmezejében szereplő dokumentumszámok közül azok fogják szolgáltatni a szelektálás kívánt eredményét, amelyek valamennyi kiválasztott fogalom lapjain szimultán fellelhetők. Az adatvisszakeresés lényeges mozzanatát tehát a dokumentumszámok kollacionálása teszi. A kollacionálás elvén működő rendszereknek a többivel szemben az a rendkívül nagy előnye, hogy információszolgáltatáskor elmarad a gyűjteménynek teljes egészében történő átválogatása.

E csoportba több eljárás, illetve az eljárásokhoz szerkesztett berendezés sorolható. A különböző berendezések mechanizáltsági foka egészen széles spektrumot ölel fel a manuális eljárásoktól az automatizáltság csúcspontján levő elektronikus szerkezetekig.

A legprimitívebb mechanizmussal az uniterm coordinate indexing dolgozik. Adatfelvevő lapjain tíz függőleges rovat van, melyek felzeteit a 0—9 számjegyek teszik. E rovatokba írandók a vonatkozó dokumentumok azonossági számai az utolsó számjegyük szerinti megoszlásban. A visszakeresés úgy történik, hogy a válasz összes karakterisztikumainak lapjait egymás mellé fektetve leolvassuk a valamennyi lapon egyöntetűen fellelhető dokumentumszámokat.

A vizuális lyukkártyák már magasabb fokú mechanizáltságot nyújtanak. Rovatok helyett jelmezőhálózattal bírnak, amelynek számjelzett metszéspontjai a dokumentumszámok egyértelmű leképezését teszik lehetővé. A dokumentumszámmal azonos hálózati pont lyukasztása rögzíti, hogy a lap felzetében szereplő fogalom a vonatkozó dokumentum descriptora. Az információszolgáltatás művelete egyszerű és gyorsabb, mint az előbbi eljárásnál. A kollacionálás automatikusan végezhető a válaszfogalmak lapjainak egymásra fektetése és átvilágítása által. Az árnyékolt környezetből fénylő pontok mutatják a keresett dokumentumok azonossági számait.

A vizuális lyukkártyáknak, mint adatfelvevő közegnek felhasználásával speciális dokumentációs gépeket is szerkesztettek (Alpha-Matrex gépek). Ezeknél a kártyák manipulálását mechanikus, illetve a fejlettebb megoldásban elektromos berendezéssel végzik. Az utóbbinál a megválaszolható téma descriptorainak elektromos írógépen való lebillentyűzésével automatikusan emelődnek ki a gyűjteményből a kollacionáláshoz szükséges lapok.

Vannak másféle adattároló közeggel működő kollacionáló dokumentációs szelektorok is. A Special Index Analyser adattároló gyűjteménye Hollerith-kártyákból áll, s a descriptoronként felfektetett kártyákba jelölik a dokumentumszámokat. A kártyákat descriptoronként elhatárolt kvantumokban, blokkokban tárolják. A válaszadás céljából végzett szelektálás úgy történik, hogy a gép a válasz descriptorainak megfelelő blokkokat külön-külön vizsgálja át. Az elsőből kiírja az összes dokumentumszámokat egy papír lyukszalagra. A második blokk dokumentumszámait kollacionálja a kiírtakkal, s az egyezőket a szalagnak egy további részére írja át. A harmadik blokk dokumentumszámait már ez utóbbiakkal kollacionálja, s az egyezőket ismét kiírja. Az utolsó kollacionálás eredményeként kiírt dokumentumszámok lesznek a keresettek, melyeket elektromos írógép olvasható jegyzékbe ír át.

A Minicard-berendezés szintén felhasználható kollacionálásos adatvizsszakeresésre, bár eredetileg dokumentumonkénti válogatásra van szerkesztve. Kisméretű filmlapokra készített adatfelvételei dokumentumonként vannak felfektetve, s a szükséges kódjelzeteken kívül még a dokumentumnak vagy tartalmi kivonatának fotokopikus másolatát is tartalmazzák. A filmlapokat kétezres-kvantumokban tárolórekeszekbe tölti a gép s így kerülnek a szelektorba, mely rekeszenként dolgozza fel az anyagot. De a berendezés a filmlapokról automatikusan másolatokat tud készíteni. Így lehetővé válik, hogy minden descriptor számára külön másolat készüljön. Ha a rekeszekbe a filmlapokat a descriptorjelzetek szerinti csoportosításban rakatják be a géppel, a szelektorral kollacionálásos információkeresés végezhető. A Minicard-berendezés tehát a kollacionálásos eljárás előnyét egyesíteni tudja saját adatfelvételeinek amaz előnyével, hogy az adatlapokon a dokumentumnak vagy tartalmi kivonatának szövegét is tárolja. Ennek következtében az információt a dokumentumnak vagy referátumának fotokópiájában tudja nyújtani.

Az elektronikus adatfeldolgozógépek kategóriájába sorolható s a kollacionálás elvén működő első kísérleti berendezés az EDIAC (Electronic Data Investigating Automatic Computer). Memóriájában 250 descriptort és 100 dokumentumszámot tárol. Külön jellegetessége, hogy a referencékérdés megválaszolása során nemcsak a vonatkozó dokumentumszámokat adja meg, hanem a megválaszolt témához kapcsolódó descriptorokat is közli. Így válaszában útmutatást ad a szóbanforgó témával kapcsolatban föltehető további érdeklődő kérdésekre.

Az újabban elterjedő RAMAC-berendezés (Random Access Memory Automatic Computer) különleges memóriaszerkezeténél fogva kiválóan felhasználható kollacionálásos adatkeresésre. A gép memóriája függőleges tengelyen centrálisan elhelyezett fémkorongokból áll. A korongokon kon-

ctrikus körökben futnak mágneses jelzetek formájában az adatfelvételek. Minden kör mentén az egy-egy descriptorhoz tartozó dokumentumszámok vannak bejelölve. A beíró-leolvasó mágnesesfej a korongok tengelyével párhuzamosan mozgatható és a kívánt descriptorok korongjaira, illetve azokon belül jelzet-köreire közvetlenül ráállítható. Az adatok bejelölése és leolvasása a korongnak a fej alatti körbefordulása folyamán történik. A dokumentumok analízisének eredményeként a dokumentumonként felfektetett adatfelvételek eredeti formájukban kerülnek a gépbe, amely automatikusan csoportosítja át a „dokumentumszám/descriptorjelzetek” típusú adatfelvételeket „descriptorjelzet/dokumentumszámok” típusúvá, hogy a kollacionálásos adatvisszakeresés alapját megteremtse. A válogatás eredményeként kapott dokumentumszámok nyomán a memóriának egy másik részében tárolt bibliográfiai gyűjteményből elektromos írógéppel komplett címleírásokat tartalmazó bibliográfiai jegyzéket állít elő.

Még magasabb nivójú adatszolgáltatást nyújtanak a nagyméretű elektronikus adatfeldolgozó berendezések (data processing systems), amelyekkel szintén végezhető kollacionálásos adatvisszanyerés. Belső memóriájuk a válogatáshoz szükséges adathalmazt tartalmazza. Az innen eredményként kapott dokumentumszámok alapján külső memóriájukból a vonatkozó dokumentumok referátumainak vagy esetleg maguknak a dokumentumoknak szövegét gyorsnyomtatógéppel közvetlenül olvasható formában prezentálják.

A szokványos lyukkártyarendszerű adatfeldolgozógépeket is alkalmazták kollacionálással működő információkereséshez, de ez eddig csupán részműveletek kivitelezésére korlátozódott. Főleg a posting-művelet — a dokumentumszámoknak a szempontok szerinti lapokra való reávezetése — mechanizálását, illetve könnyítését és gyorsítását próbálták elérni ily módon.¹ De olyan megoldások is találhatóak, amelyekben magát a kollacionálást lehetővé tevő és előkészítő műveleteket végzik lyukkártyagépekkel.²

Indokoltnak látszott tehát vizsgálat alá vonni azt a problémát, hogy a szokványos Hollerith-gépeket minő eljárással lehet önálló és teljesérvényű kollacionálásos információkereséshez felhasználni.

A lyukkártyagépek a numerikus adatokat egyszerűbb műveletekkel, gyorsabban és könnyebben tudják feldolgozni, mint az alfabetikusokat vagy alfanumerikusokat. Ezért az irodalomfeltárás során előálló mindkét adatféleséget: a descriptor-t és a dokumentum azonosítási jelzetét tisztán numerikus formában kell a Hollerith-kártyán szerepeltetni. A legközvetlenebbül adódó mód a természetes számsor egymásután következő számainak igénybevétele. Kétféle jelzetet egymástól megkülönböztetni. Például úgy, hogy a dokumentumszámok öt-, a descriptorjelzetek négy-számjegyű számok legyenek.

A descriptoroknak számjelzettel való helyettesítése természetesen megkívánja kettős segédnyilvántartás vezetését. Az egyikben alfabetikus rendben a descriptorok sorakoznak, mellettük számjelzeteik. A másikban a számjelzetek következnek természetes rendjükben, feltüntetve descriptor-ekvivalensüket. Minden újonnan felmerülő descriptor-nak ki kell jelölni számjelzetét, s mindkét nyilvántartásba fel kell venni.

A szempont szerint felfektetett irodalomfeltárási adatlapok eredeti jellege azt kíváná, hogy a Hollerith-kártyákon is a szokásos elrendezés érvé-

nyesüljön: az egyik almezőben a descriptor jelzete szerepeljen, a másikban a vonatkozó dokumentumszámok. Ez az elrendezés azonban itt nem alkalmazható. Ha ugyanis egy Hollerith-kártyán a dokumentumszámok almezőjében akár megnevezett, akár nevezetlen zónákban több dokumentumszámot szerepeltetünk, akkor a rendelkezésre álló lyukkártyagépekkel nem tudjuk végrehajtani a különböző kártyákon szereplő, előttünk ismeretlen dokumentumszámok egyeztetését. Ezért a szokásostól eltérő adatbejelölési módot kell választani, hogy a dokumentumszámok kollacionálásának kívánalmát a gépek képességeivel összhangba hozhassuk. Ez pedig abban áll, hogy a Hollerith-kártyán felvett két almező mindegyikébe csak egy-egy jelzet kerül a szokásos Hollerith-bejelöléssel, az egyikbe a descriptor jelzete, a másikba a descriptorral jellemzett dokumentum — de csak egyetlen dokumentum — azonossági száma. Ez az elrendezés képezi tulajdonképpen a probléma megoldásának kulcsát. (Ábra 131. lapon)

A dokumentumok analízise nyomán készített primér adatfelvétel természetesen dokumentumonként van felfektetve. Tartalmazza a dokumentum azonossági számát, valamint descriptorait. A descriptorok mellé odavezetendő numerikus jelzeteik a már említett segédnyilvántartásból. Az így elkészített primér adatfelvételi lapokról készülnek a lyukkártyák. Mégpedig minden dokumentumhoz annyi, ahány descriptor-tadtunk neki. A lyukkártyák elkészítése rendkívül egyszerű. Mindegyikükbe rendre egy másik descriptor-jelzete kerül, míg a dokumentumszám valamennyiöknél ugyanaz lesz. Mivel a kártyák jelmezejének csak kisebb hányada van így igénybevéve, a kártyafelület üresen maradó nagyobbik részébe, amely mintegy 70 oszlopot ölelhet fel, a dokumentum rövidített címlírása és fontosabb bibliográfiai adatai jelölhetők be. (Ld.: az ábrát.) Így annak a hátránynak kompenzálásaként, hogy minden descriptorjelzettel és minden dokumentumszámmal külön kártya készül, azt nyerjük, hogy az információszolgáltatás a dokumentumoknak nemcsak azonossági számait, hanem címadatait is nyújthatja. Ehhez természetesen az szükséges, hogy a kártyába belyukasztott adatok a kártya peremére betűkre átírtan legyenek felvezetve. Ez a lyukasztással egyidejűleg végezhető, ha a kártyákat az ún. írólyukasztógépen készítik. A közönséges lyukasztógépek bejelöléseit az interpreter géppel kell olvasható formára átírni.

A kész lyukkártyák gyűjteménnyé kumulálódnak, mely háromféle módon tárolható: teljesen rendezetten, részlegesen rendezetten vagy rendezetlenül. Teljesmértvű a gyűjtemény rendezettsége, ha a kártyák a descriptorjelzetek szerinti csoportokba vannak gyűjtve és ezeken belül a dokumentumszámok sorrendjében következnek. Részleges a rendezettség, ha a descriptorok szerinti csoportképzés megvan ugyan, de azokon belül a kártyák elhelyezése esetleges. Rendezetlen pedig a gyűjtemény akkor, ha a kártyák teljesen keverten, azaz mind a descriptorjelzetre, mind pedig a dokumentumszámmra való tekintet nélkül következnek egymásután.

A kártyagyűjtemény eme különböző rendezettségi állapotai egyaránt jelentenek előnyt és képeznek hátrányt. Minél alacsonyabb fokú a rendezettség, annál kevesebb a kártyáknak a gyűjteménybe való reponálásával járó munka. Az újonnan elkészült kártyákat a gyűjteménybe be kell helyezni, az információszolgáltatás céljából kisselektáltakat pedig vissza kell tenni.

Ez a két művelet, mivel nem tetemes mennyiségekről van szó, manualiter végezhető. Ennyi kártyának besorolása érdekében ugyanis nem érdemes a tekintetbe jöhető gépeket igénybe venni, illetve azokon egy jelentős kártya-kvantumot átfuttatni csak azért, hogy annak egymástól távol fekvő helyeire egy-egy vagy legfőljebb néhány kártya kerüljön.

A kártyák reponálása szempontjából a háromféle rendezettség a következőképpen nyilvánul. A rendezetlenül tárolt gyűjteményhez a kártyákat egyszerűen hozzátesszük, az egymásutániságra való tekintet nélkül. Ha részleges a rendezettség, akkor a besorolandó kártyamennyiség előbb a descriptorjelzetek szerint előrendezendő. Aztán meg kell keresni minden előforduló descriptorjelzet csoportját, amelyekhez a kártyák simán csatolhatók. Ha azonban a gyűjtemény teljesen rendezett, akkor az előrendezett kvantumból a kártyákat a descriptoraiknak megfelelő csoportokban a dokumentumszámuknak megfelelő helyre kell szabályszerűen besorolni.

Közbevetőleg megjegyezzük, hogy az „egy descriptorjelzet/egy dokumentumszám” típusú adattárolás éppen határesetet képez a descriptorok szerinti és a dokumentumok szerinti adattárolási módok között. Az ily kártyagyűjteményt éppúgy fel lehet használni a descriptorszerinti adattárolásból kiinduló kollacionálásos információkereséshez, mint a dokumentum-szerinti adattárolásból kiinduló ún. felkutatásos információvisszanyeréshez. A határeseti jellegből adódik, hogy gyűjteményünket a másik vezérellem — a dokumentumszám — szerinti rendezettség szempontjából is vizsgálhatnók. Ezt azonban mellőzzük, mivel az ez esetben alkalmazható adatkeresőeljárások már nem a kollacionálás elvén alapulnának.

A gyűjtemény rendezettségi foka az információszolgáltatás céljából végzett gépi műveletekkel viszont fordított arányban áll. Minél kevésbé rendezett a gyűjtemény, annál több rétvűbb gépi művelettel érhető el az eredményt szolgáltató kollacionálás. Ezért a rendezettség három fokának megfelelőleg háromféle gépi eljárás szerkeszthető.

1. A *rendezetlen gyűjtemény* tekinthető a legáltalánosabb esetnek, mivel ilyenkor a válogatásra kerülő gyűjteményt illetőleg semmiféle előfeltétel vagy megkötöttség nem állván fenn a gépi műveleteket a feladat teljességének megoldására kell szerkeszteni.

A megoldás megszerkesztésénél abból kell kiindulnunk, hogy két parallel adathalmazunk van: a descriptorjelzetek és a dokumentumszámok halmaza. Mindkét halmazban teljesen rendszertelenül következnek a tagok egymás után; s mindkettőben egyes tagok többszörösen is előfordulnak, amiben azonban szintén nem nyilvánul semmiféle szabályszerűség. Az egyedileg megkülönböztethető tagok száma ennél fogva mindkét halmazban kisebb a tagok összes számánál. Mindkét halmaznak egy-egy tagja a közös tároló közeg (kártya) folytán egymáshoz van kapcsolva, adatkárokat képeznek. A párok tagjai között azonban semmiféle olyan reláció nem áll fenn, hogy az egyikből a másikra bárminemű következtetést vagy utalást lehetne meríteni. Az első halmaz minden tagjának van tehát a másik halmazban egy hozzátartozója, mely vele közös kártyán szerepel. Ha azonban az első halmaz egy olyan egyedi tagját tekintjük, amely abban többször fellép, akkor ennek a másik halmazban annyi hozzátartozója lesz, ahányszoros az ismétlődés multipllicitása. Hogy ezek a hozzátartozók melyek és a másik halmazban hol

találhatók, azt csak az az első halmaz tagjainak és a pár-relációnak átvizsgálásával lehet megállapítani.

Problémánk ezek után így fogalmazható: melyek a második halmaznak ama tagjai, amelyeket az első halmaz bizonyos kiválasztott tagjainak szimultán előfordulása jellemez. A megoldás több lépésből tevődik össze. Az első lépésben az első halmaz ismeretlen egymásutánjából meghatározott tagokat kell kiemelnünk. A reátalálás a kiválasztott tagok jegyeinek és a halmaz egymást követő tagjai jegyeinek összevetése és az egyezés tényének megállapítása nyomán következik be. E művelet eredményeként az első halmazból egy olyan részhalmaz fog elhatárolódni, amely felöleli valamennyi kiválasztott tagot mégpedig teljes multiplicitásukkal. De ezáltal egyúttal a második halmazból is leválasztunk egy részhalmazt, amelyben viszont az előbbi részhalmaz összes tagjainak valamennyi hozzátartozója benne foglaltatik. Az így kapott két részhalmaz természetesen szintén rendezetlen. Nyilvánvaló, hogy következőleg azt kell megállapítani, hogy a második részhalmaznak melyek ama egyedileg megkülönböztethető tagjai, amelyeknek a saját részhalmazukban nyilvánuló előfordulási multiplicitása egyezik az első részhalmazhoz tartozó tagok számával. A második lépésnek tehát egy szétválogató műveletnek kell lennie, amelynek célja, hogy a második részhalmaz egymásutánjából kikeresse az abban fellelhető — előttünk ismeretlen — egyedileg megkülönböztethető tagokat, s azokat csoportokba gyűjtse. Ezt aztán egy kollacionáló műveletnek kell követnie, melynek során a csoportok állományait összevetve az határozandó meg, hogy melyek a második részhalmaznak ama tagjai, amelyek valamennyi csoportban egyaránt szerepelnek. Ezek fogják képezni a keresett eredményt.

A megoldás elvi részének tisztázása után lássuk a tényleges kivitelezést. A gépi műveleteket a descriptor jelzetek almezején végzett válogatással kezdjük. A gyűjtemény rendezetlen állapotára tekintettel a szorter géppel végeztetjük a szelektálást.³ Mégpedig a reference-kérdés alkatelemeit képező valamennyi descriptorjelzet tekintetbe vételével egyetlen rendezőválogató gépműveletben. Ennek befejezése után a kiszelektált kártyákat a válogatás alapjául szolgáló descriptorjelzetek szerint külön-külön csoportokban kapjuk, de az egyes csoportok rendezetlenek lesznek, azaz a csoportokon belül a kártyák nem fogják követni a dokumentumszámok sorrendjét. E művelettel a gyűjteményből kiemelődnek mindama kártyák, amelyeken a megválasztandó téma alkatelemeit képező descriptorok valamelyike jelzet formájában rajta van. A csoportok kvantumai természetesen nemcsak a keresett kártyákat fogják tartalmazni, hanem több-kevesebb más kártyát is. A következő részfeladat tehát a csoportok kártyaanyagának szétválasztása keresettekre és nemkeresettekre. Mivel azonban nem tudjuk azt, hogy melyek a keresett és nemkeresett kártyák, a szétválasztást úgy kell végrehajtanunk, hogy az egyúttal a keresettség jellegének rögzítését is nyújtsa. Először is a dokumentumszámok szerint sorbarendezzük a csoportok kártyamennyiségeit. Ezt a szorter géppel végeztetjük a dokumentumszámok almezején végrehajtott rendezéssel. Mivel a csoportok a teljes gyűjteménynek még együttesen is csak törtrészt tartalmazták, a művelet viszonylag rövid idő alatt végrehajtható, habár a szükséges gépfutamok száma egyenlő a csoportok számának és a dokumentumszám számjegyei számának

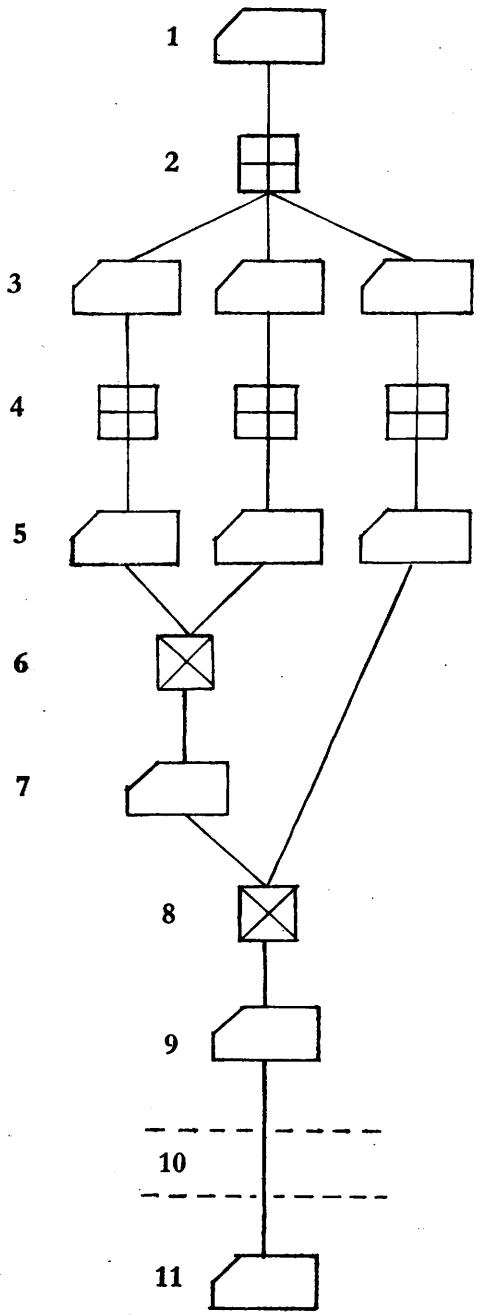
szorzatával, tehát aránylag elég magas. Arra kell vigyázni, hogy az egyes csoportok kártyakvantumai a rendezés után is elkülönítettek maradjanak. A sorrendberakás után következik a csoportok anyagának kollacionálása, amit a kollatorgéppel végeztetünk. Kiválasztjuk a legnagyobb és legkisebb csoportot. Az előbbi szelektálandó kártyamennyiségként, az utóbbit pedig keresőkártyákként alkalmazva a kollatorral többkártyás válogatást hajtunk végre. Ennek során a gép a keresőkártyákba lyukkasztott dokumentumszámok alapján vizsgálja át a szelektálandó mennyiséget, s abból kiemeli mindazokat a kártyákat, amelyeknek ekvivalensei a keresőkártyák között előfordultak. E művelettel tehát a nagyobbik csoportból leválasztjuk ama kártyákat, amelyeknek dokumentumszámai a két csoport mindegyikében megtalálhatók voltak. Mármost az így kapott kiszelektált mennyiséget keresőkártyákként alkalmazva hasonló műveletet hajtunk végre egy harmadik csoporton. A most kiválogatott kártyák már az eddig vizsgált mindhárom csoportnak közös kártyái. Az eljárást ugyanígy — azaz a mindenkor kiszelektált kártyákat keresőkártyaként és a soronkövetkező csoportot szelektálandó mennyiségként véve — folytatjuk az összes többi csoportokon. Nyilvánvaló, hogy a művelet befejeztére az utolsó csoportból kiszelektált kártyák fogják tartalmazni ama dokumentumszámokat, amelyek az összes csoportokban egyaránt fellelhetők. Ezek lesznek a választ tartalmazó dokumentumok.

Az eljárás menetét részletesen feltüntető folyamatdiagram (ld. 136. lapon) a következő:

1. kártyagyűjtemény (rendezetlen);
2. szorter szelektálást hajt végre 1-en a megválaszolendő téma descriptorjelzetei szerint;
3. kiszelektált kártyák descriptorjelzetenként elhatárolt csoportokban;
4. szorter a csoportok kártyáit dokumentumszámok szerint sorbarendezi;
5. rendezett állományú csoportok;
6. kollator többkártyás szelektálást hajt végre a dokumentumszámok almezején, az első két csoport igénybevételével;
7. első kollacionáló menetben kiszelektált kártyák;
8. kollator 7-et keresőkártyaként alkalmazva többkártyás válogatást végez az 5-nek harmadik csoportján;
9. második kollacionáló menetben kiszelektált kártyák;
10. kollator lefolytatja a további kollacionálásokat az 5-nek többi csoportjain, a 8 analógiájára;
11. utolsó kollacionáló menetben kapott kártyák. Végeredmény.

Ez a megoldás, melyet a rendezetlenség esetére, mint a legáltalánosabb jellegű állapotra adtunk, a problémának egyetemes megoldása. Amint a rendezetlen gyűjteményből egy bizonyos fokon végrehajtott rendezéssel részlegesen rendezett gyűjteményt kapunk, amelyből egy másik fokon megnyilvánuló rendezés teljesen rendezett gyűjteményt tud létesíteni, úgy az egyetemes megoldásból bizonyos bevezető lépések elhagyásával a részlegesen rendezett gyűjtemény információkereső eljárását állíthatjuk elő. Majd ebből további lépések elmaradása folytán a teljesen rendezett gyűjtemény számára kapjuk meg a megoldást.

2. *A részlegesen rendezett gyűjtemény a rendezetlentől abban különbözik,*



hogyan az állomány a descriptorjelzetek szerint csoportokban van tárolva. Ez az állapot hasonlít ahhoz a helyzethez, ami akkor áll elő, amikor a rendezetlen gyűjteményből a megválaszolendő téma descriptorainak megfelelő kártyákat szorter géppel kiválogatva csoportokba gyűjtve kapjuk azokat. Az eltérés abban nyilvánul, hogy a részlegesen rendezett gyűjtemény egészében ilyen csoportokból áll, míg az előbbi esetben csak néhány csoporttal volt dolgunk.

Információkeresés esetén most tehát azzal kezdődik az eljárás, hogy a gyűjteményből a referencékérdés descriptorainak megfelelő csoportokat manuálisan kiemeljük. Ezeket aztán végrehajtjuk mindama gépi műveleteket, amelyek a diagram 4-ik és következő lépéseiben vannak feltüntetve. E művelet sor fogja adni problémánk megoldását a részlegesen rendezettség esetére.

3. Ha a gyűjtemény teljesen rendezett az egyetemes megoldás műveleti lépéseiben még tovább mehetünk. A teljes rendezettség ugyanis, mint az egyetemes megoldás egy fázisa a diagram 5-ik lépésénél áll elő. Ott természetesen a teljes rendezés is csak ama néhány csoportra vonatkozik, amelyeket a rendezetlen gyűjteményből a referencia szolgáltatás céljaira kiemeltünk. A teljesen rendezett gyűjtemény azonban összességében ily állapotban van. Az információszolgáltató eljárás a teljesen rendezett gyűjteményeknél is azzal kezdődik, hogy a descriptorjelzetek szerinti egymásutánban álló csoportok közül kézzel kiemeljük az informálandó téma descriptorainak megfelelőket. Ezeknek gépi kimunkálása ugyanazon műveletekkel történik, mint amelyeket a diagram 6-ik és következő lépései tüntetnek fel.

A standard IBM-gépeknek kollacionálásos információvisszanyeréshez való igénybevételének problémáját mások is vizsgálták.⁴ Módszerük tulajdonképpen azonos azzal, melyet itt a teljes rendezettség esetére ismertettünk. A probléma megoldását ők tehát csak a gépi műveletek szempontjából legegyszerűbb esetre adták meg.

Tárgyalásunk teljessége érdekében még egy különleges esetre is ki kell térnünk. Ez akkor áll elő, ha a megválasztandó téma egyetlen descriptorral jellemezhető, tehát a szelektálást csak ezen egyetlen descriptor szerint kell végezni. A részlegesen vagy teljesen rendezett gyűjteményeknél ilyenkor egyszerűen a vonatkozó descriptorjelzetnek kártyacsoportja fogja a referenciaanyagot szolgáltatni, tehát ezt a csoportot kell kézzel kiemelni a gyűjteményből. A rendezetlen gyűjteményeknél azonban gépi eljárásra van szükség, mert a kiválasztott descriptor jelzetével bíró, s a gyűjteményben elszórt, összes kártyákat ki kell emelni. E munka kollatorral vagy szorterrel végezhető.

A kollatorral végzendő szelektáláshoz egy keresőkártyát kell készíteni, melybe a válogatás alapjául szolgáló descriptorjelzet van bejelölve ugyanazon oszlopokból álló zónába, mint a gyűjtemény kártyáin. A gép e kártya alapján a gyűjtemény átfutó tömegéből kiválogatja a kívántakat. A művelethez a teljes kártyatömeg egyszeri átfuttatása szükséges.

Szorter géppel kétféleképpen hajtható végre a válogatás. Vagy egy rendező-válogató művelettel, amelyet a kiválasztott descriptorjelzet szerint folytatunk le. A gyűjtemény állományának egyszeri teljes, valamint a descriptorjelzet második és további számjegyeinek megfelelő rendező-menetekben ismételt, de rohamosan csökkenő kvantumú átfuttatását igényli.

Ha azonban rendelkezésre áll a szorterre felszerelhető többszlopos szelektor, akkor ezt a descriptorjelzet számjegyeinek oszlopaira állítva egyetlen menetben kiemelhetők a szóbanforgó jelzettel bíró összes kártyák.

Befejezésül még annyit, hogy az elmondottakban ismertetett eljárások úgy is tekinthetők, mint a közismert uniterm coordinate indexingnek lyuk-kártyagépekkel kivitelezett teljesérvényű mechanizált változatai.

Jegyzetek

1. Sanford, J. A. — Theriault, F. R.: *Problems in the application of uniterm coordinate indexing*. College and Res. Libr. 17 (1956). 1. 19—23. l. Gull, C. D.: *Posting for the uniterm system of coordinate indexing*. American Doc. 7. (1956.) 1. 9—21. l. Weinstein, S. H. — Drozda, R. J.: *Adaptation of coordinate indexing system to a general literature and patent file: machine posting*. American Doc. 10 (1959). 2, 122—129. l.
2. Peakes, G. L.: *The unit card system in the indexing of internal technical reports*. Progress report in chemical literature retrieval. New York. 1957. 149—164. l. Peakes, G. L.: *Experience with the unit card system for report indexing*. Information systems in documentation. New York. 1957. 306—327. l. Whaley, F. R.: *A deep index for internal technical reports*. Information systems in documentation. New York. 1957. 352—383. l.
3. Leírását l. *Magy. Könyvszle.* 76 (1960). 1. 16—17. l.
4. Lancaster, E. R.: *An information searching system based on the IBM Collator*. IBM Technical Newsletter 1956. jún. 12, 12. s. köv. l. Faden, B. R.: *Information retrieval: punched card techniques and special equipment*. Special Libr. 50 (1959). 6. 244—249. l.