

OROSZ GÁBOR

Információkereső eljárások a nálunk használatos Hollerith-gépekhez

A hazai dokumentáció területén kezdenek megjelenni a gépi információkereső eljárások alkalmazása iránti igények. Ezek kielégítésének legközvetlenebbül adódó módja a nálunk használatos Hollerith-gépek igénybevétele. Ennek azonban előfeltétele megfelelő használati eljárások kidolgozása, mert az információkeresés feladatai és igényei a szokásos gépi adatfeldolgozó munkától (statisztika, könyvelés) teljesen eltérő jellegűek. Az alábbiakban egy ilyen dokumentációs adatfeltáró és visszakereső rendszert ismertetünk, illetve két eljárást annak a nálunk meglévő gépekkel történő kivitelezésére.

A gépi megmunkálás céljaira a dokumentumokat (könyvek, folyóirat-cikkek stb.) alkalmasan fel kell dolgozni. Ez azzal kezdődik, hogy a dokumentum tartalmi átvizsgálása alapján megfelelő tárgyszószerű fogalmakat választanak témájának megjelölésére. E jellemzőket (descriptorokat) a gépi feldolgozás számára jelzetekre (kódokra) kell áttenni. A jelzetek elvileg állhatnak betű- vagy számjegyekből, vagy mindkettőből. A gyakorlatban a kizárólag számjegyekből álló jelzetek használata az elterjedtebb, mert ezek gépi feldolgozása könnyebb. A számjelzeteket vagy valamilyen osztályozó rendszerből (például az ETO-ból) szokták venni, s ekkor kiválasztásuk alapja a descriptoroknak a rendszer szerinti osztályozása, vagy pedig egy elhatárolt számhalmazból (pl. a 0000—9999 számok közül), amikor viszont megválasztásuk tetszés szerint történik (hozzáértve természetesen azt a feltételt, hogy valamely fogalom jelölésére egyszer már igénybe vett jelzetet más fogalom számára felhasználni nem szabad). A jelzetek felvételére dokumentumként egy vagy több Hollerith-kártyát fektetnek fel, amelyeknek megfelelően elhatárolt részeibe, az ún. tárgyi almezőkbe jelölik be azokat. Az almezők (a szokványos és nálunk is meglévő Hollerith-gépek esetében) a kártyán függőlegesen vannak elhelyezve, és a kártya egymás mellett fekvő számoszlopainak meghatározott csoportját foglalják magukba. Az almezők mindegyikébe egy-egy jelzet kerül. Bejelölésének módja a szokásos Hollerith-jelölésmód, azaz a jelzetet alkotó számjegyek helyi értékének megfelelő oszlopban az alaki értéknek megfelelő jelhelyre tett lyukasztás. A műveletet a külön e célra szolgáló lyukasztógép billentyűnyomásra végzi (az írógép klaviatúrájához hasonló működéssel).

A tárgyi almezőkön kívül föltétlenül helyet kell biztosítani a kártya jelmezében a dokumentum azonossági számát felvevő zónának, hogy a kártya és a vonatkozó dokumentum közötti kapcsolat egyértelműleg rögzítve legyen. Felvehető természetesen a kártyára a dokumentummal kapcsolatos más jellegű adatok is.

A tárgyi almezők lehetnek megnevezettek vagy nevezetlenek. A megnevezett almezőkbe csak az almezőkkel kapcsolatban valamilyen módon

determinált jelzetek kerülhetnek. Ez az eset például, ha a kártyán 10 darab 0—9 számokkal megnevezett almezőt vesznek fel, melyekbe az ETO-ból vett számjelzeteket jelölik főcsoportonkénti megoszlás szerint (1. ábra). Az ilyen rendszernek súlyos hátránya, hogy a vonatkozó jelzetcsoportoknak (szakcsoportoknak) csak egy-egy jelzete vihető a megfelelő almezőbe. Ha valamely csoportból még egy jelzet szerepeltetése szükséges, már egy második kártyát kell felfektetni, még akkor is, ha az első kártya többi tárgyi almezői üresen is maradnának. Ez viszont részint a válogatásra kerülő kártyagyűjteményt duzzasztja nemkívánatos mértékben, részint a szimultán több szempont szerinti szelektálásnál teremt nehéz problémákat. Ha azonban a tárgyi almezők nevezetlenné, akkor bármelyik jelzet bármelyik almezőbe tehető, teljesen tetszés szerinti választás nyomán (2. ábra). Az almezők igénybevételekor a jelzeteknek számsorbeli rendje vagy akár a leképezett fogalmaknak értelmi egymásutánja teljesen közömbös. A jelzeteknek és az almezőknek függetlensége egymástól e rendszer kiváló előnye. Így ugyanis a jelzetek (s ezektől visszafelé haladva: a fogalmak) megválasztását illetőleg semmiféle korlátozás nem áll fenn. A dokumentumból annyi témajellemző fogalmat emelhetünk ki, ahány almezőt tudunk biztosítani a kártyán. Azt tehát így — leszámítva a mellőzhetetlen azonossági szám részére fönntartott néhány oszlopos sávot — teljes egészében a tárgyi feltárást szolgáló jelzetek felvételére lehet felhasználni. Ez pedig elegendővé teszi dokumentumonként egyetlen kártya felfektetését. A 80 oszlopos kártyán ugyanis:

6 számjegyből álló jelzetek esetén 12 tárgyi almező + 8 oszlop az azonossági számok részére ;

5 számjegyből álló jelzetek esetén 14 tárgyi almező + 10 oszlop az azonossági számok részére ;

4 számjegyből álló jelzetek esetén 18 tárgyi almező + 8 oszlop az azonossági számok részére ;

3 számjegyből álló jelzetek esetén 24 tárgyi almező + 8 oszlop az azonossági számok részére határolható el. Ekkora jelzetmennyiség felvételére más rendszereknél 2—3 vagy még több kártya szükséges, ami nem csupán kártyafogyasztásban jelent többletet, hanem a szelektálás időtartamát nyújtja kellemetlenül hosszúra.

A dokumentációs adatfeltárást másik felét az információadás céljából végzett visszakeresés teszi, melynek az a rendeltetése, hogy az érdeklődő kérdésekre válaszanyagot tartalmazó dokumentumok azonossági számait megállapítsa. Az azonossági számok kikeresése pedig az információk kérdés jelzetre áttett alkatelemeinek (fogalmainak) és a tárgyi feltárást során dokumentumonként felfektetett jelzeteknek koincidienciáján alapszik. A válogatást végző gép a kártyagyűjteménynek átfuttatása folyamán az egyezésnek esetenkénti bekövetkezését érzékeli, és a vonatkozó kártyát a gyűjteményből kiemeli.

Ez a szelektálási eljárás is teljesen eltérő a kétféle rendszerben. Az információk kérdés descriptorainak jelzeteit az egyik esetben előre megállapítható almezőben kell megkeresni, a másikban viszont valamelyik almezőben kell megtalálni. E művelet szempontjából tehát a megnevezett almezőrendszer az egyszerűbb. Ilyenkor a gép érzékelő szerkezete a jelzet szakcsoportjának megfelelő almezőre eleve ráállítható, s a gépnek nem kell egyebet tennie, mint figyelni az átfutó kártyák vonatkozó almezőjében levő jelzeteknek egy keresőkártyába lyukasztott jelzettel való egyezését. Ha ellenben a szelektálás

alapjául szolgáló jelzet — föltételezve, hogy egyetlen jelzet szerinti válogatást végzünk — az almezők bármelyikében lehet, s egyik kártyán itt, a másikon ott, a harmadikon amott van, az összes almezőket végig kell pásztázni, ha e jelzettel ellátott kártyák valamennyiét ki akarjuk emelni. Még sokkal komplikáltabb lesz a helyzet akkor, ha több jelzetről álló konfigurációt kell keresni (pedig ez a gyakrabban előálló eset). A konfigurációk tárgyi jelentése ugyanis nem változik, bármilyen is bennük az elemek sorrendje. Tehát n jelzet esetén $n!$ -féle jelzetsorrendű konfiguráció lehetséges, amelyek mindegyikét ki kell válogatni. Ezek pedig az m számú almezőben ($m \geq n$)

$$\frac{m!}{(m-n)!}$$

-féleképpen helyezkedhetnek el. Hogy mekkora számok lépnek itt fel, azt mutatja a következő táblázat:

3 jelzet esetén a konfigurációk száma	6
4 jelzet esetén a konfigurációk száma	24
5 jelzet esetén a konfigurációk száma	120
6 jelzet esetén a konfigurációk száma	720

Egy 14 almezős kártyát véve, ennek almezőiben ez a

6 konfiguráció	2 184-féleképpen helyezhető el
24 konfiguráció	24 024-féleképpen helyezhető el
120 konfiguráció	240 240-féleképpen helyezhető el.

Tehát még nagy kártyatömegek esetén is könnyen megtörténhetik, hogy nem akad két olyan kártya, amelyben a keresett jelzetek mind azonos almezőkben feküdnének.

A megnevezetlen almezős rendszert eredetileg az IBM 101-es típusú statisztikai adatfeldolgozó géphez szerkesztette E. GARFIELD dokumentáció-technikai kutató.* E modern gépnek magasrendű képességei fölvetették benne a Hollerith-kártya olyan gazdaságos kihasználásának lehetőségét, amelyre korábban a szokásos Hollerith-gépek használata mellett nem is gondoltak. A 101-es géppel a szabad almezőkön is kivitelezhetővé vált a multidimenzionális szelektálás. De a szokásos lyukkártyagépekhez sem azelőtt, sem azóta nem használták e rendszert. Ennek oka abban rejlik, hogy e gépekre nem volt megoldva gyakorlatilag elfogadható alkalmazásának módszere. Az az eljárási mód ugyanis, mely a kollátorgép (mint a szokásos Hollerith-gépek közül a dokumentációs szelektálásra legalkalmasabb) képességeinek tekintetbevételével közvetlenül adódik, bár elvileg helyes, gyakorlatilag hasznavehetetlen, mert először az első jelzet szerinti szelektálást kellene végrehajtani az összes almezőkön, azután megismételni ugyanezt a második jelzet szerint, majd a harmadik szerint, és így tovább. Ez azonban a teljes kártyatömegnek $m \times n$ átfuttatását jelentené (pl. 14 almező és 4 jelzetről álló konfiguráció esetén a teljes kártyagyűjtemény 56-szoros átválogatását), ami mind időtartamban, mind pedig gépigénybevételben túl van a reális lehetőségek legszélsőbb határán is.

* GARFIELD, E.: *Preliminary report on the mechanical analysis of information by use of the 101 Statistical Punched Card Machine*. Amer. Document. 5 (1954) 1, 7—12. l.

A rendszer előnyös volta ily előzmények ellenére mégis arra indított, hogy megvizsgáljuk, szerkeszthető-e olyan eljárás, mely gyakorlatilag elfogadható lehetőséget nyújtana a szokásos lyukkártyagépekkel végrehajtott multidimenzionális szelektálásra. Sikerült is a probléma megoldására három eljárást kidolgozni, melyek közül kettőt az alábbiakban ismertetünk. A megoldásban az információkereső műveleteknél eddig nem használt módszert alkalmazunk. A szelektálást — melyet kizárólag a tárgyi jelzetek alapján szoktak végezni — megosztjuk. Első részét a dokumentációs törzskártyák gyűjteményén a tárgyi jelzetek szerint folytatjuk le, másik részében pedig a kiszelektált törzskártyákról készült másodkártyák azonossági számait válogatjuk át. Az eljárások megszerkesztésénél azt a célt tartottuk legfontosabbnak, hogy a teljes kártyatömeg átfuttatásainak számát a tárgyi almezők számára szállítsuk le (pl. a fentebb említett példában 56-ról 14-re). Így ugyanis az eljárások a gyakorlati használhatóság ésszerű korlátait nem lépik át.

Eljárásaink ismertetésénél a 2. ábrán bemutatott kártyaelrendezést vesszük alapul. A dokumentum azonossági számának felvételére elhatárolt rész 10 számoszlop szélességű, a tárgyi almezők 5 számjegyből álló jelzetek felvételére vannak méretezve. A jelzetek a 00,000—99,999 számok közül választhatók. A jelzésrendszer kapacitása tehát kerekén 100 000 jelzet, ami túl bőven elegendő, és gyakorlatilag ki sem mérhető. Mindaz azonban, amit a következőkben elmondunk, természetesen más jelzetterjedelmű, illetve kapacitású jelzésrendszerek esetében is érvényes.

A válogatás mindkét eljárásban a rendezőgéppel (sorter) történik. Mivel az eljárások legfontosabb mozzanata a szelektálás művelete, ezt behatóbban ismertetjük. A sortergéppel végzett szelektálás tulajdonképpen a szokásos rendezési eljárás, amely megfelelő manuális közreműködéssel válogatást eredményező műveletté válik. A válogatási eljárás megértéséhez tehát ismerni kell a sorter rendező eljárásának lefolyását. — A könyvtárosok sokat gyakorolják a jegyenkénti (betű- vagy számjegyenkénti) kártyarendezést. Módszere abban áll, hogy a kevert kártyatömegeből egyenkint leemeljük a lapokat és a rendezés alapjául szolgáló jelkép (szó- vagy számkép) baloldaltól legelső helyen álló jegyei szerint csoportokba gyűjtjük. Így az első betűjegy, illetve a legmagasabb helyi értékű számjegy szerint rendezzük. A kapott csoportokat azután hasonló eljárással a második helyen álló jegy nyomán válogatjuk szét másodcsoportokba, és így tovább mindaddig, amíg az utolsó rendű csoportokig el nem jutunk. A sortergép rendező művelete elvileg ugyanígy folyik le, csak a jegyek egymásutánjában szoktak jobbról balfelé haladni. Az eljárást a rendezendő zónában levő számok legalacsonyabb helyi értékénél kezdik, s a teljes kártyatömegnek ismételt átfuttatásával gépmenetenként rendre egy-egy soron következő magasabb helyi értéket véve folytatják le. A gyakorlati kivitelezés a következőképpen történik. Az első menetben a gép az egyesek helyén álló számjegyek leképező lyukasztásainak érzékelése nyomán a megfelelő gyűjtőrekeszekbe irányítja az egyes kártyákat. Így az eredetileg kevert kártyatömeget az egyesek helyén álló számjegyeknek megfelelően tíz csoportba osztja. A géprekeszekből a kártyakvantumokat a számok alaki értékének egymásutánjában kell kiszedni. Legelőször a 0-ás rekeszből kell kivenni a lapokat, oly helyzetben hagyva, amint abban feküdtek; azután ezekre reá kell helyezni az 1-es rekesz kártyáit, emezekre a 2-es rekeszben találtakat. Legfelülre a 9-es rekesz kártyái kerülnek. Ezt a kártyatömeget

sorrendjének gondos megtartása mellett újra a gépbe kell adagolni, mely, mivel közben érzékelő szerkezete a zóna balról második oszlopára lett állítva, azt a következő menetben hasonló eljárással ez oszlop lyukasztásai nyomán rendezi ismét tíz csoportba. E második menet után a kártyák a zóna számainak utolsó két számjegye szerint vannak szigorú számsorba rendezve. Hasonló módon történik a következő, egyre magasabb helyi értékű számjegyek szerinti rendezés is. A zóna utolsó oszlopán (balról legelső oszlop) végzett rendezés után a teljes kártyatömeg a lapjain szereplő számok emelkedő egymásutánjában fog sorakozni. Az azonos számú kártyák természetesen egymás mellett lesznek.

A sorternek ez a rendezési eljárása szelektáló műveletté tehető azáltal, hogy a második és minden további menetben már nem a teljes kártyagyűjtemény, hanem csupán a szelektálás alapjául választott szám(ok) vonatkozó jegyeinek kártyakvantumai kerülnek csoportokba osztás céljából a gépbe. Ehhez természetesen emberi közreműködés szükséges. Ugyanis az egyes menetek után csak a kiválasztott szám vonatkozó jegye rekeszének anyagát kell a következő átfuttatáshoz felhasználni, a többi rekeszek anyaga, mint fölösleges, félreteendő. Az utolsó menet után a kiválasztott szám legmagasabb helyi értékű jegyének megfelelő rekeszben lesznek találhatóak a kiszelektált kártyák. — Világítsuk meg ezt egy példával. Legyen feladatunk egy kevert tömegből mindama kártyáknak kiválogatása, amelyeknek a 22—26 oszlopokat felölelő zónájában a 60914 kulcsszám szerepel. A művelet a 26-ik oszlopon végzett rendezéssel kezdődik. E menet után csak a 4-es gyűjtőrekeszben talált kártyamennyiséget tartjuk meg, a többi rekeszek tartalmát érdektelenként félreteesszük. A második menetben már csak ezt a töredék kártyamennyiséget rendeztetjük a 25-ik oszlop bejelölései alapján, s végeztével csak az 1-es rekeszből kikerülő kártyamennyiséget tartjuk meg további feldolgozásra. Ebben már csak olyan lapok lesznek, melyeknek kulcsszámai 14-gyel végződnek. A 24-ik oszlopon végzett harmadik rendezési menet után a 9-es rekesznek, a 23-ik oszlopon végrehajtott negyedik átfuttatás után pedig a 0-ásnak kártyakvantuma kerül további rendezésre. Az utolsó, azaz a 22-ik oszlopon végzett rendezés folyamán a 6-os rekeszbe gyűjtődnek mindazon lapok, amelyeken a kiválasztott 60914-es kulcsszám szerepel. Tehát ezeket sikerült a kevert tömegből kiválogatni.

Az eljárás lefolytatható természetesen több kiválasztott kulcsszám tekintetbe vételével is, azaz ily módon többdimenziós szelektálás végezhető. Ekkor azonban az első menet után valamennyi kiválasztott szám utolsó számjegyeinek megfelelő rekeszek anyagát kell — egymástól külön tartva — félretenni a második menet számára, s e menettől kezdve az egyes kulcsszámok kártyakvantumait külön-külön almenetekben kell a vonatkozó oszlopon végzett rendezésekkor átfuttatni, nehogy esetleg az egyik kulcsszám kártyái a másikéival keveredjenek. — Megjegyezzük, hogy a szelektálási művelet lefolytatható a legmagasabb helyi értéktől az alacsonyabbak felé haladó rendezésekkel is, teljesen analóg módon.

A sorter géppel tehát ilyen — hogy úgy mondjuk — szelektáló eljárással közvetve tudunk szelektálást végrehajtani. A művelethez szükséges a teljes kártyatömegnek egyszeri átfuttatása, az első menetben. A további meneteknél már csak rendre csökkenő, mégpedig erősen csökkenő kártyamennyiségek szerepelnek. Nagy előnye az eljárásnak, hogy előzetes rendezésre nincs szükség, mert ezúttal tulajdonképpen maga a rendezés a szelektálás.

A többi gépi műveletet, amelyekre az eljárások folyamán sor kerül, saját helyükön ismertetjük oly mértékben, amennyire a folyamatok összefüggéseinek megértéséhez szükséges.

Az *első eljárás* alapgondolata az, hogy ha az almezőnként lefolytatott szelektálások alkalmával kieső kártyákból rendre kijegyezzük a vonatkozó dokumentumok azonossági számait, akkor ezek között lesznek egyszer, kétszer, ... n -szer előfordulók. Nyilvánvaló, hogy csupán az n -szer ismétlődő azonossági számú dokumentumok fognak az információs kérdésre válaszanyagot tartalmazni. Ezek ugyanis ama kártyák azonossági számai, melyek valamennyi keresési jelzet nyomán szelektálódtak ki az m almező közül valamelyik n különbözőn. Problémánk megoldását a sok azonossági szám közül eme n -szeres ismétlődésűeknek kiemelése fogja adni.

Az eljárás gépi menete az első almezőn a sorterrel végrehajtott szelektálással kezdődik. A kiválogatott kártyákról a kártyamásológéppel (doppler) másolatokat készítünk. E másodkártyákba csak az azonossági szám lesz átmásolva (3. ábra). A másolatkészítés után a kiszelektált törzskártyákat egyszerűen hozzátesszük a gyűjteményhez, s következhetik a második almezőn végzendő szelektálás a sortergéppel. Ennek eredményéről szintén készítünk másodkártyákat. Ugyanígy járunk el a további almezőkön is. Az utolsó almezőn kiszelektált kártyák másolatainak elkészítése után együtt van az összes almezőn végzett szelektálások során kiesett valamennyi kártya azonossági számainak egyenként külön-külön másodkártyákba átvitt speciális kollekciója. Ezt a kártyamennyiséget aztán sorterrel szigorú számsorba rendeztetjük. A kapott egymásutánban ugyanazon azonossági számmal 1 vagy 2, vagy ... n kártya szerepel. Amely azonossági számmal n kártya található egymás mögött, az egy keresett dokumentum száma. Ha a másodkártyák össz mennyisége csekély (néhányszor tíz darab), akkor az n -szer ismétlődő számú kártyák könnyen kiválogathatók kézzel is. Bőséges szelektálási eredmény esetén a manuális válogatás már nem célszerű: hosszú időt igénylő és fárasztó munka. Géppel kellene végeztetni. Viszont gépeinkkel nem hajtható végre olyan válogatás, amellyel az ismeretlen számoknál fellépő n -szeres ismétlődésű kártyákat ki lehetne emelni.

A nehézséget egy műfogással sikerült áthidalni. A kollátorral végezhető egy speciális művelet: a csoportelső kártya leválasztása. Az emelkedő számsorba rendezett kártyahalmazból, amelyben az azonos kulcsszámmal bíró kártyák csoportokat alkotnak, a kártyatömeg átfutása során a gép minden csoport első kártyáját leválasztja és külön gyűjti. Ha ezt a műveletet a másodkártyák előzetesen számsorba rendezett (tehát csoportokba rakott) tömegén $n-1$ -szer lefolytatjuk, kiemelődnek a mindenkor csoportok első kártyái, s az utolsó menet után visszamaradó kártyamennyiségben már csak azok a kártyák fognak szerepelni, amelyek a művelet kezdetekor az n darabból álló csoportok utolsó kártyái voltak. Így sikerül kiszűrni a másodkártyák kollekciójából az n -szeres előfordulású azonossági számokat. Vagyis megkapjuk azon dokumentumok azonossági számait, amelyeknek dokumentációs kártyáin a tudakozó kérdés valamennyi elemének jelzete szerepel, tehát az információkeresés eredményét szolgáltatják.

Arra az esetre, ha kollátorgép nem áll rendelkezésre, egy másik módszert is közlünk, mellyel a végeredményt képező azonossági számok szintén megállapíthatók. Ehhez a táblázógép igénybevétele szükséges. A másodkártyáknak sorterrel emelkedő számsorba rendezett kollekcióját, amelyben tehát az



3. ábra
Másodkártya. Bejelölve a 49,317.59 azonossági szám

ugyanolyan azonossági számmal bíró lapok egymás után következnek, a tabellálógépen úgy futtatjuk keresztül, hogy a gép minden azonossági számot első előfordulásakor írjon ki, s azután számolja meg, hogy hány kártya fut keresztül a gépen a vonatkozó azonossági számmal, és ezt a számot is írja ki. Így egy jegyzéket kapunk, amely a másodkártyákon fellépő valamennyi azonossági számot tartalmazza, s közli egyúttal azt is, hogy a vonatkozó azonossági számmal hány kártya található a másodkártyák kollekciónjában. Nyilvánvaló, hogy a legmagasabb előfordulási számhoz tartozó azonossági számok a keresettek. Ezeket a jegyzék átnézése során abban manualiter kell kijelölni, vagy abból kiírni.

Az edmondottak megértésének megkönnyítésére közöljük az eljárás folyamatdiagramját is, amely szemléletesen mutatja be a műveletek menetét.

Fölvehető az eljárás hátrányaként az a kényszerűség, hogy minden információkereséshez nem csekély mennyiségű másodkártyát kell készíteni, amelyeknek pedig csak egyetlen esetben van szerepük. Ez kétségtelenül így van, de a gyakorlatban e tény nem oly előnytelen, mint amilyennek első látszatra mutatkozik. Fennáll ugyanis a másodkártyák ismételt igénybevételének lehetősége. A korábbi másodkártyák későbbi információkereséseknél újólag felhasználhatók olyképpen, hogy az újabb műveletek során adódó azonossági számokat a kártyák további zónáiba jelöljük be. Ha 10 oszlopos azonossági zónákat alkalmazunk, akkor a másodkártyák nyolc különböző kereséshez használhatók, s így a kártyafogyasztás lényegesen csökken.

A *másik eljárás* kidolgozására éppen az a törekvés vezetett, hogy a másodkártyák számát a minimumra csökkentjük. Amíg az előbbi eljárásnál valamennyi szelektálás során kieső összes kártyákról készülnek másolatok, addig ez utóbbinál csupán a szelektálás nyomán jelentkező dokumentumonként. A kiválogatott kártyákon levő azonossági számokról másolat csak a vonatkozó törzskártyának első kiszelektálása alkalmával készül, s az egyes válogatási menetek során való kieséseit e másodkártyán megfelelő jelöléssel rögzítjük. E célból a másodkártyákon az azonossági számok zónáján kívül még egy függőleges zónát veszünk, amely annyi számoszlopot ölel fel, ahány keresési jelzetünk van, azaz n -et. Minden jelzet számára, számsorbeli egymásutánjuk rendjében, kijelölünk egy oszlopot. Amikor a törzskártya valamely jelzet szerinti válogatáskor kiesik, ezt másodkártyáján a megfelelő oszlop egy kiválasztott jelhelyén lyukasztással — kiesési jelzéssel — jelezzük. Így a másodkártyákon az eljárás befejeztére azt is rögzítjük, hogy törzskártyáik végeredményben hányszor szelektálódtak ki. Most tehát a másodkártyák kollekciónjában minden előforduló azonossági számmal csak egyetlen kártya fog szerepelni, de e lapokon 1 vagy 2 vagy . . . n kiesést jelző lyukasztás lesz. Az információ tudakozódás megválaszolását az n -szeres kieséssel jelölt kártyák azonossági számai fogják nyújtani (4. ábra).

A gépi műveleteket az előkészítő munkálatok után, éppúgy mint az előbbi eljárásnál, a sorterrel végzett szelektálással kezdjük az első almezőn. A kiválogatott kártyákról, melyeket a keresési jelzetek szerinti csoportokban kapunk a szelektáló művelet végén a sortergéből, a dopplerrel most is másodkártyákat készítünk a dokumentum azonossági számával. A csoportokban álló másodkártyákat minden csoport saját oszlopának X jelhelyén lyukasztással jelöljük (kiesési jelzés), majd az így elkészített másodkártyákat sorterrel egyetlen emelkedő számsorba rendeztetjük. A kiválogatott törzskártyákat pedig esatoljuk a gyűjteményhez, s következhetik a második almezőn vég-

zendő szelektálás. Az ennek során kiválogatott és a keresési jelzeteknek megfelelő csoportokban álló törzskártyákkal most azonban más műveleteket végzünk. Legelőször is a kollátorgéppel összekerestetjük (párosítjuk) a kiszelektált törzskártyák és az első menet folyamán készült másodkártyák közül az ugyanolyan azonossági számmal bírókat. E műveletet azért végezzük, hogy a kiszelektált kártyák csoportjaiból kiemeljük azokat a lapokat, amelyeknek azonossági számáról már készült másodkártya az első almezőn végzett válogatást követőleg. A párokból aztán a kiesést jelölő lyukasztás nyomán sortergéppel ún. splittszelektáló menetben leválasztjuk a másodkártyákat. A másodkártyáknak így előálló csoportjaiba bejelöljük a vonatkozó szelektálási jelzeteknek megfelelő újabb kiesési jelet, az X lyukasztást, az erre kijelölt oszlopon. E kártyákon most már kétféle kiesési jelzés lesz, ami azt jelenti, hogy az ezeknek a másodkártyáknak megfelelő törzskártyák a keresési jelzetek közül kettőt tartalmaznak. A kiválogatott törzskártyáknak a párosításkor visszamaradt példányai, amelyekhez a másodkártyák közül pár nem került, e második szelektálás során estek ki első ízben. Ezekről ugyanúgy, mint fentebb az első almezőn végzett válogatás után, csoportonként másodkártyákat készítünk, amelyeket szintén megfelelő kiesési jelzéssel látunk el. Ezután az összes másodkártyákat — azaz az első menetben készültek közül mind azokat, amelyek második kiesési jelzést kaptak, mind pedig azokat, amelyek ilyen nem kaptak, továbbá a most készülteket — egybegyűjtjük és a sorterrel egyetlen emelkedő számsorba rakatjuk. A kiszelektált törzskártyákat pedig hozzátesszük a gyűjteményhez, s végeztetjük a harmadik almező szerinti válogatást. Ennek, valamint az összes többieknek menete azonos a másodikéval. Csupán az $m-n$ -ediket követő menetekben van annyi változás, hogy az ekkor első ízben kiszelektált azonossági számokról a másodkártyák készítése elmarad. Ezekre a másodkártyákra ugyanis n kiesési jelzés már nem kerülhet, tehát ezek a végeredmény kialakulásában nem játszhatnak szerepet.

Az utolsó almezőn végzett válogatás és a csatlakozó műveletek elvégzése után együtt van a másodkártyák kollekcója. Ebből kell azokat a kártyákat kiemelni, amelyeken n kiesési jel található, vagyis e zónának valamenynyire oszlopában van X lyukasztás. Ez a művelet könnyen elvégezhető éppen az n darab X lyukasztás alapján végrehajtott válogatással. Mégpedig vagy a kollátorral, az e célra készült keresőkártya segítségével, vagy a sorterrel végzett n számú splittszelektáló menettel, — vagy kézi válogatással, kevés kártya esetén. E legutolsó lépésben kapott kártyák azonossági számai szolgáltatják az információs kérdés megválaszolásához szükséges dokumentumokat.

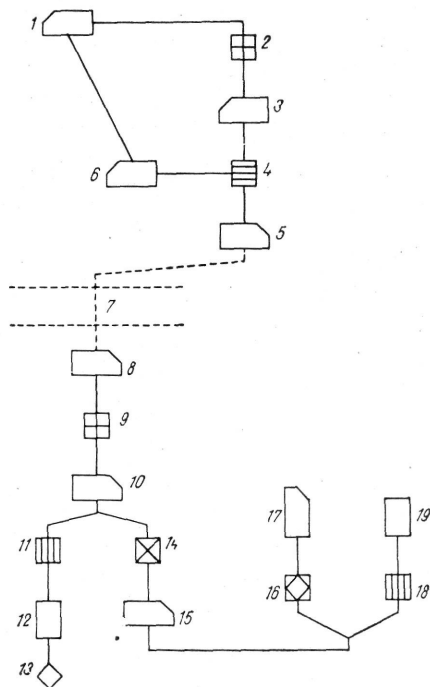
A kártyafogyasztás csökkentése második eljárásunknál természetesen más téren követelt engedményeket, s így ez az eljárás az előbbihez képest több és különbözőbb jellegű lépésből tevődött össze. Ez azonban használhatóságát nem befolyásolja. Megfelelő szervezéssel (ami minden Hollerith-munkának amúgy is előfeltétele) ezen — így elmondva komplikálnak tűnő — eljárás a logikusan egymásba kapcsolódó szakaszok praktikus rendezett folyamat adja. E folyam időtartama a valóságban sokkal rövidebb annál, amit az elmagyarázás nyomán az olvasó feltételez, mert a különböző gépeken végzendő munkálatok jó része szimultán futtatható, s ez az eljárás teljes idejét jelentősen csökkenti. A másodkártyák ismételt felhasználásának lehetősége itt is fennáll, bár az újbóli igénybevételek száma most kevesebb lesz, mert esetenként két zóna terjedelmét kell lefoglalni a kártyán. Az eljárás menetét és rész munkálatait feltüntető folyamatdiagramot szintén közöljük.

Amint az elmondottakból láthattuk, a végeredményt mindkét eljárásban a másodkártyákba lyukasztásokkal bejelölt azonossági számok képezik. Ezeknek olvashatóvá tételére két mód is van. Vagy a felíratógéppel írhatjuk át a lyukasztásokat számok formájában a kártya felső peremére, vagy a tabellálógéppel jegyzéket készítettünk az azonossági számokról. Gyakorlott egyének azonban közvetlenül is le tudják olvasni a lyukasztások nyomán a megfelelő számjegyeket.

A *harmadik eljárás* elvi felépítése nagyjából a másodikat követi. Az eltérés az, hogy a kiesési jelzések alkalmazása most az almezők, illetve a szelektálási menetek egymásutánjában történik. (Az első almezőn végzett szelektálás nyomán készült másodkártyák az első oszlopban kapnak X jelet, a második menet során készültek a második oszlopban stb.) Ily kivitelezéssel azonban az eljárás folyamatának szerkezete többértévé válik.

Az eljárások alkalmazhatóságának határát, illetve azon kártyagyűjteménynek terjedelmét, amelyhez gyakorlatilag elfogadható időtényezővel használhatók, a sortergép sebessége szabja meg. A legidőigényesebb művelet ugyanis a szelektálás, mert almezőnként egy-egy ízben a teljes kártyamennyiség átfuttatását kívánja. A sortergépek óránkénti teljesítménye a típusoktól függőleg 24—60 000 kártya. Ezt véve alapul, a gyűjtemény nagyságrendje 10^4 -ben állapítható meg. A 100 000-en felüli dokumentum-mennyiségekhez már elektronikus berendezés szükséges.

Adott esetben az eljárások bármelyikének alkalmazására kerül is sor, kétségtelen, hogy akár az elsőnek nagyobb kártyafogyasztása, akár a másodiknak összetettebb menete mind elmaradnak ama, a dokumentációs feltárás mélységében és az információszolgáltatás differenciáltságában jelentkező előny mellett, amit a nevezetlen almezők rendszerének alkalmazása nyújt.



Első eljárás
folyamatdiagramja

1. Dokumentációs törzskártyák gyűjteménye
2. Sortergép 1-et az első almezőn szelektálja
3. 1-ből kiválogatott törzskártyák
4. Dopplergép 3-ról másolatokat készít az azonossági számokkal
5. Elkészült másodkártyák
6. A másodkártyák elkészülte után a kiszelektált törzskártyák visszakerülnek a gyűjteménybe
7. Szelektálás a többi almezőkön az 1—6 lépések analógiájára
8. Összes almezők átszelektálása után előálló másodkártyakollekció
9. Sorter emelkedő számsorba rendezi (csoportokba rakja) 8-at
10. Másodkártyák rendezett kollekciója

*innen két irányban folytatódhat az eljárás;
egyik alternatíva*

11. Tabelláló gép 10-ből kiírja az azonossági számokat és a vonatkozó kártyák multiplícitási számait
12. Táblázat, mely az összes azonossági számokat tartalmazza
13. n-szeres multiplicitású azonossági számoknak emberi művelettel való kikeresése 12-ből

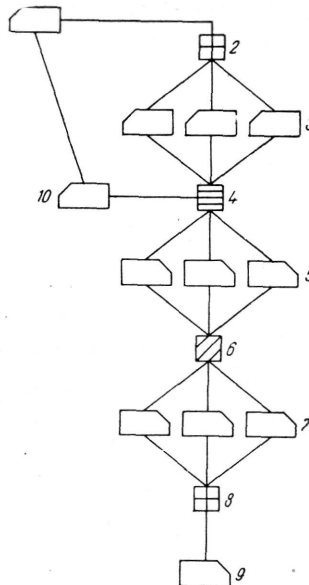
másik alternatíva

14. Kollátor n-1-szer csoportelső-kártya leválasztását végzi 10-en
15. 14 után visszamaradó másodkártyák, melyek a keresett azonossági számokat tartalmazzák

A lyukasztással bejelölt azonossági számokat kétféle módon lehet olvashatóvá tenni:

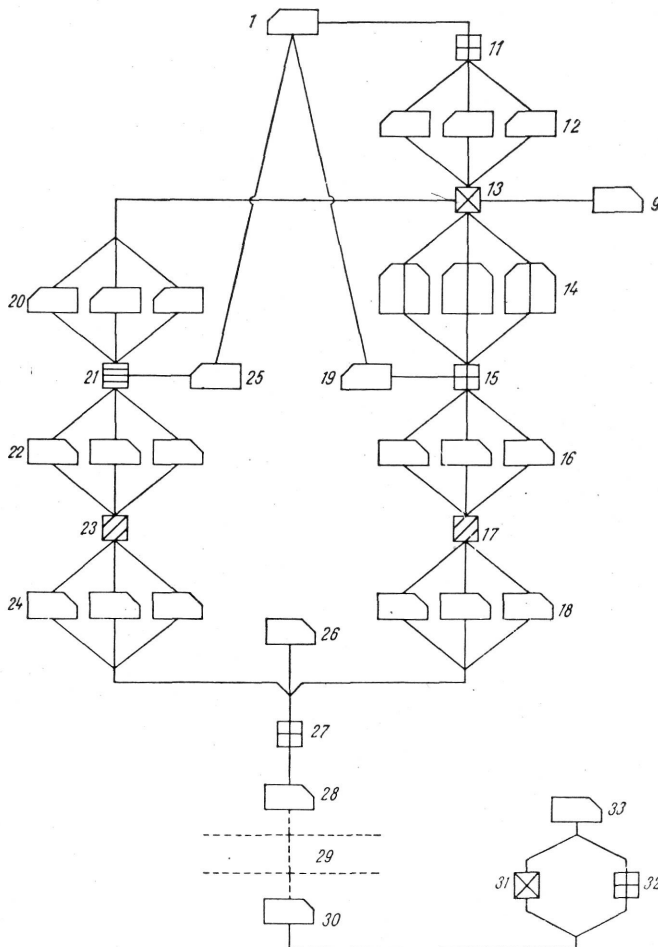
16. Feliratozó gép a 15 kártyák felső peremére írja át a beléjük lyukasztott azonosság i számokat
17. Végeredményt szolgáltató kártyák a feliratozott azonossági számokkal
vagy
18. Tabelláló gép a 15 kártyákból táblázatosan írja ki az azonossági számokat
19. Végeredményt adó táblázat

*Második eljárás folyamatdiagramja (1. rész)
Az első almezővel kapcsolatos műveletek*



1. Dokumentációs törzskártyák gyűjteménye
2. Sortergép 1-et az első almezőn szelektálja
3. 1-ből kiválogatott törzskártyák a keresési jelzeteknek megfelelő csoportokban
4. Dopplergép 3-ról csoportonként másolatokat készít az azonossági számokkal
5. Elkészült másodkártyák csoportjai
6. Lyukasztógép 5-be csoportonként belyukasztja a megfelelő kiesési jelzéseket
7. Kiesési jelzésekkel ellátott másodkártyák csoportjai
8. Sortergép 7-et egyetlen emelkedő számsorba rendezi
9. Első almezőn végzett szelektálás nyomán készült másodkártyák összessége
10. 4 után a kiszelektált törzskártyák visszahelyeződnek a gyűjteménybe

Második eljárás folyamatdiagramja (2. rész)
 A második és további almezőkkel kapcsolatos műveletek. Végműveletek



1. Dokumentációs törzskártyák gyűjteménye
11. Sortergép 1-et a második almezőn szelektálja
12. Kiszelektált törzskártyák a keresési jelzeteknek megfelelő csoportokban
13. Kollátorgép párosítja 12-ből és 9-ből az ugyanolyan azonossági számú kártyákat
14. Kártyapárok a keresési jelzeteknek megfelelő csoportokban

15. Sorter a 14 csoportokból splittszelektálással leválasztja a másodkártyákat
16. Leválasztott másodkártyák csoportjai
17. Lyukasztógép 16-ba csoportonként belyukasztja az újabb kiesési jelzéseket
18. Két-két kiesési jelzéssel ellátott másodkártyák csoportjai
19. 15 után a kiszelektált törzskártyák közül azok, amelyek a párokban szerepeltek, visszakerülnek a gyűjteménybe
20. 13 után pár nélkül maradt kiszelektált törzskártyák csoportjai
21. Dopplergép 20-ról csoportonként másolatokat készít az azonosságai számokkal
22. Most készült másodkártyák csoportjai
23. Lyukasztógép 22-be csoportonként bejelöli a kiesési jelzéseket
24. Egy kiesési jelzéssel ellátott másodkártyák csoportjai
25. 21 után az ott szerepelt törzskártyák visszatétnek a gyűjteménybe
26. 9-ből visszamaradt azon másodkártyák, amelyekhez a 13 során pár nem került
27. Sortergép a 18-at, 24-et és 26-ot egyetlen emelkedő számsorba rendezi
28. A másodkártyák összessége a második almezőn végzett szelektálás után
29. Szelektálás a többi almezőknél a 11—28 műveletek analógiájára. Az $m-n$ -ediket követő almezőknél a 20 analogonja visszakerül a törzskártyák gyűjteményébe, s a 21—25 lépések elmaradnak
30. Összes almezők átszelektálása után előálló másodkártyakollekció, melyben különböző számosságú kiesési jelzéseket tartalmazó kártyák vannak

A 30-ból az n kiesési jelzést tartalmazó kártyák kiemelhetők

31. vagy kollátorgéppel, az n kiesési jelzésnek megfelelő lyukasztású keresőkártyával
32. vagy sortergéppel, splittszelektálással
33. a keresett azonosságai számokat tartalmazó másodkártyák.

A 33 lapjain lyukasztással bejelölt azonosságai számok olvashatóvá tétele ugyanúgy történik, mint az első eljárásnál (első folyamatdiagram, 16—19 műveletek).

**GÁBOR OROSZ: ZWEI VERFAHREN DES INFORMATIONSAUFFINDENS
VERMITTELS HOLLERITH-MASCHINEN**

Bei der Anwendung der Hollerith-Maschinen für das Informationsauffinden lässt sich die Hollerith-Karte auf mancherlei Weise benutzen. Die eine jener ist die Methode der ungenannten Unterfelder (Abb. 2). Dabei kann die Karte — abschlägich die für die Identitätsnummer des Dokumentes begrenzte Zone — voll und ganz für die sachlichen Unterfelder in Anspruch genommen werden. In die sachlichen Unterfelder werden die Schlüssel der das Thema des Dokumentes charakterisierenden Begriffe einmarkiert.

Die Anwendung der ungenannten Unterfelder ist äusserst rationell, da sie auch unter Anlegung einer einzigen Karte je Dokument die Aufnahme der Schlüssel von bedeutender Anzahl ermöglicht.

Das System der ungenannten Unterfelder wurde ursprünglich für die IBM Statistikmaschine vom Typ 101 entwickelt. Mit Hilfe dieser Maschine war die Selektion nach mehreren Gesichtspunkten auch im Falle der ungenannten Unterfelder möglich. Für die üblichen Hollerith-Maschinen wurde jedoch dieses System nicht angewendet, da seine praktische Anwendung nicht gelöst werden war.

Der äusserst vorteilhafter Charakter des Systems veranlasste uns doch zu untersuchen, ob ein Verfahren, das die praktisch annehmbare Durchführung der mit den üblichen Lochkartenmaschinen vollzogenen multidimensionalen Selektion bieten würde, konstruiert werden kann. Es gelang uns zur Lösung des Problems drei Verfahren auszuarbeiten, deren zwei wir an dieser Stelle darlegen möchten. Bei der Entwicklung der Verfahren war unser Hauptzweck, dass die Anzahl der Durchläufe der vollen Kartenmenge auf die Anzahl der sachlichen Unterfelder reduziert worden sei.

Die Auswahl erfolgt im Falle beider Verfahren vermittelt der Sortiermaschine. Die mit der Sortiermaschine durchgeführte Selektion ist eigentlich das übliche Sortierverfahren, das eine Auswahl ergebende Operation wird dadurch, dass es in dem zweiten und jedem weiteren Kartendurchlauf schon nicht die völlige Sammlung, sondern nur die Kartenmengen der betreffenden Ziffern der zum Grund der Selektion gewählten Schlüssel zwecks Sortierung in weitere Gruppen in die Maschine gelangen. Das Verfahren kann auch unter Berücksichtigung mehrerer Schlüssel vollzogen werden, d. h. wir sind solcherweise imstande, eine multidimensionale Selektion durchzuführen.

Der Grundgedanke des ersten Verfahrens ist, dass es, wenn wir aus den bei den je Unterfeld durchgeführten Selektionen ausgefallenen Karten die Identitätsnummern der betreffenden Dokumente nacheinander ausschreiben unter denselben einmal, zweimal . . . n-mal Vorkommende gibt. Es ist offensichtlich, dass nur die Dokumente von n-mal sich wiederholenden Identitätsnummern ein Antwortmaterial auf die Informationsfrage enthalten werden.

Der maschinelle Vorgang des Verfahrens beginnt mit der am ersten Unterfeld vermittelt der Sortiermaschine durchgeführten Selektion. Über die aussortierten Karten werden mit Hilfe des Kartendopplers Kopien gefertigt. In diese Zweitkarten werden nur die Identitätsnummern überschrieben (Abb. 3). Nach der Fertigstellung der Kopien geben wir die aussortierten Stammkarten einfach der Sammlung zu, und es kann die am zweiten Unterfeld durchzuführende Selektion folgen. Nach der Fertigstellung der Kopien der am letzten Unterfeld aussortierten Karten ist die Kollektion der Zweitkarten beisammen. Dann wird diese Kartenmenge vermittelt der Sortiermaschine in eine steigende Zahlenreihe geordnet. In dem erhaltenen Nacheinander kommen mit gleicher Identitätsnummer 1 oder 2 oder . . . n Karten vor. Die Identitätsnummer, unter der n Karten nacheinander zu finden sind, ist nun die Nummer eines gesuchten Dokumentes. Es sind also diese Identitätsnummern von n Multiplizitäten irgendwie zu erhalten. Dies kann auf zwei Wege erreicht werden: mit Hilfe entweder der Kollator- oder der Tabelliermaschine. Vermittels der Kollatormaschine kann man eine spezielle Operation durchführen: die Abtrennung der gruppenersten Karten. Von der in steigende Zahlenreihe geordneten Kartenmenge, in der die Karten von gleicher Schlüsselzahl Gruppen bilden, trennt die Maschine während des Durchlaufes der Kartenmenge die erste Karte jeder Gruppe automatisch ab und sammelt dieselben gesondert. Wenn diese Operation an der geordneten Kollektion der Zweitkarten n—1-mal vollzogen wird, heben sich die ersten Karten der jeweiligen Gruppen heraus, und in der nach dem letzten Gang remanenten Kartenmenge kommen nur die Karten vor, die zum Beginn der Operation die letzten Karten der aus n Stücken bestehenden Gruppen waren. Wir können aber das Endergebnis auch mit Hilfe der Tabelliermaschine erhalten. Die geordnete Kollektion der Zweitkarten lassen wir durch die Tabellier-Maschine laufen, derart dass die Maschine jede Identitätsnummer bei ihrem ersten Vorkommen ausschreibt und zählt dann, wie viele Karten mit betreffender Identitätsnummer durch die Maschine läuft und sie schreibt diese Nummer gleichfalls aus. Somit erhalten wir eine Liste, die alle, in den Zweitkarten figurierenden Identitätsnummern enthält und zugleich mitteilt, wie viele Karten mit betreffender Identitätsnummer in der Kollektion der Zweitkarten auffindbar sind. Es ist offensichtlich, dass es die zu der höchsten Vorkommensnummer gehörenden Identitätsnummern sind, die wir gesucht hatten.

Zur Ausarbeitung des anderen Verfahrens wurden wir gerade durch die Strebung geführt, dass die Zahl der Zweitkarten auf das Minimum vermindert werde. Während Kopien bei dem früheren Verfahren über alle, im Laufe der Selektion ausfallenden Karten gefertigt werden, geschieht es bei diesem letzteren nur je Dokument, das sich im Laufe der Selektion zeigt. Kopien über die Identitätsnummern der aussortierten Stammkarten werden nur gelegentlich der ersten Auswahl gefertigt, und ihre Ausfälle während der einzelnen Selektionsvorgänge werden auf dieser Zweitkarte mit einem entsprechenden Ausfallzeichen gemarkt. Zu diesem Zweck nehmen wir ausser der Zone der Identitätsnummern auf den Zweitkarten noch eine vertikale Zone, die so viele Spalten umfasst, wie Suchschlüssel wir besitzen, d. h. n. Für jeden Schlüssel bestimmen wir in der Ordnung des Nacheinanders in Zahlenreihe eine Spalte. Fällt die Stammkarte bei der Sortierung gemäss irgendeinem Schlüssel aus, so wird es auf ihrer Zweitkarte an einer ausgewählten Lochstelle der entsprechenden Spalte mit Lochung angezeigt. Somit wird es auf den Zweitkarten mit Beendigung des Verfahrens auch festgelegt, wievielmals ihre Stammkarten schliesslich selektiert worden sind (Abb. 4). Die Beantwortung der Informationsfrage wird nun durch die Identitätsnummern der das n Zeichen aufweisenden Karten gegeben werden.

Wir beginnen die maschinellen Operationen auch jetzt mit der durch die Sortiermaschine durchgeführten Auswahl an dem ersten Unterfelde. Über die selektierten Karten, die wir aus der Maschine in Gruppen gemäss den Suchschlüsseln erhalten, stellen wir Zweitkarten mit den Identitätsnummern der Dokumente her. Die sich in Gruppen befindlichen Zweitkarten werden an der Lochstelle X der eigenen Spalte jeder Gruppe gelocht (Ausfallzeichen). Mit den am zweiten Unterfelde selektierten Stammkarten führen wir aber schon andere Operationen aus. Zuerst lassen wir die Kollatormaschine von den selektierten Stammkarten und den im Laufe des ersten Ganges gefertigten Zweitkarten jene, die gleiche Identitätsnummern haben, zusammensuchen (paaren). Dies tun wir mit dem Zwecke, damit jene von den selektierten Stammkarten, über deren Identitätsnummern Kopien in dem ersten Selektionsvorgang bereits gefertigt worden sind, herausgehoben werden. Dann trennen wir die Zweitkarten von den Paaren ab und tragen in dieselben das neue Ausfallzeichen ein. Über die Stammkarten aber, zu denen kein Gegenstück gelangte, werden jetzt Zweitkarten hergestellt und diese ebenfalls mit entsprechendem Ausfallzeichen versehen. Der Vorgang der an den übrigen Unterfeldern durchzuführenden Selektion ist dem der zweiten gleich. Es gibt nur in den dem m - n -ten nachfolgenden Durchläufen die Änderung, dass die Fertigung der Zweitkarten über die dann zum erstenmal aussortierten Identitätsnummern unterbleibt. Nach der an dem letzten Unterfeld vollzogenen Auswahl und der Durchführung der anschliessenden Operationen ist die Kollektion der Zweitkarten beisammen. Es sind aus dieser die Karten herauszuheben, auf denen das n Ausfallzeichen vorkommt. Diese Operation lässt sich unschwer durchführen, gerade mit der auf Grund der n Stück X Lochungen vollzogenen Selektion. Und zwar entweder vermittels der Kollatormaschine — bei der Hilfe der zu diesem Zwecke gefertigten Suchkarte, oder mit den n Selektionsvorgängen der Sortiermaschine.

Die Grenze der Anwendbarkeit der zwei Verfahren, bzw. der Umfang der Kartensammlung, für die sie mit einem praktisch annehmbaren Zeitfaktor angewendet werden können, ist durch die Geschwindigkeit der Sortiermaschine bestimmt. Die Leistung der Sortiermaschinen beträgt, abhängig von den Typen, 24,000—60,000 Karten in der Stunde. Auf Grund dessen lässt sich die Grössenordnung der Kartensammlung bei 10^4 festsetzen.