

STATUS information storage and retrieval system linked with the Philips CD ROM reader. = 9th International Online Information Meeting, London, 3–5 December 1985. Learned Information, Oxford, UK and New Jersey, USA, 1985. p. 477–481.

[VII] SUMMIT, R. K.: Message from the President: information retrieval and the CD ROM. = Chronolog, 13. köt. 12. sz. 1985. p. 85:237–85:238./

(Válas György)

1985. december

### Nagy adatbázis optikai lemezen: a MEGADOC rendszer egyik első alkalmazása

Ez a dolgozat az 5. Nemzetközi Online Információs Találkozón 1981-ben tartott két előadás [1, 2] folytatása. Akkor az optikai lemezes adatbázis terveit ismertettük. Most már installáltuk a rendszert, és a vele végzett munka első tapasztalatairól számolhatunk be.

#### A Gruner und Jahr sajtóadatbankja

A Gruner und Jahr (a továbbiakban G+J) a Német Szövetségi Köztársaság egyik vezető kiadóvállalata. Dokumentációs osztálya archív szöveges információs anyagokkal és képanyaggal látja el a vállalathoz tartozó szerkesztőségek újságíróit. Legfontosabb felhasználói a *Stern*, a *Brigitte*, a *GEO*, a *Capital* és az *Impulse* magazinok, valamint a *Die Zeit* című hetilap újságírói.

A dokumentációs osztály újságkivágás-gyűjteményét több mint 200 bel- és külföldi sajtótermék cikkeiből válogatjuk. Évente mintegy 150 000 cikket indexelünk és viszünk be a gyűjteménybe, amelynek a mérete már elérte az 1,5 millió dokumentumot. A G+J speciális kereső programrendszert dolgozott ki az adatbázisban való keresés céljára *PDS* (*Press Databank Software* = sajtóadatbank szoftver) néven. Ez 1983-ban váltotta fel az 1974 óta működő hagyományos keresőrendszert. A PDS-ben megtalálhatók a korszerű keresőrendszerek szokásos eszközei: a csonkolás, a maszkolás, az adott adatmezőre korlátozott keresés, a szavak között tetszés szerinti távolságot megengedő helyzeti (proximity) operátorok. Téma és időszak szerint lehet keresni vele. A sajtóadatbank különleges igényeit szolgálja a fonetikus keresési lehetőség, amelynek segítségével a nevek a nemzetközi sajtóban szokásos

legkülönbözőbb írósmód-változatok ellenére kikereshetők.

A G+J sajtóadatbank hibrid rendszer: adatbázis, ugyanakkor primerdokumentum-szolgáltató. A keresés eredményeként a felhasználók nem a kikeresett dokumentumokról az adatbázisban tárolt információt kapják meg, hanem a kikeresett elsődleges dokumentumoknak a másolatát. Éppen ezért mindaddig kritikus pont volt az elsődleges dokumentumok gyors elérése és másolása.

#### A hagyományos módszer

Az újságkivágásokat eddig mikrofilmlapokon tároltuk. A mikrofilmlapokat különleges táregységekben helyeztük el, amelyek a kicsinyítés mértékétől függően 80 000, illetve 200 000 oldal gyors elérését tették lehetővé. 17 ilyen tárolóegységet használtunk. A régebbi anyagokat azonban hagyományos tárolószekrényekbe tettük át, azokra így nem biztosítottuk a közvetlen, gyors elérést. Ennek a rendszernek van néhány súlyos fogyatéka:

- ◆ Tízhavonként új tárolóegységet kell beállítanunk, így folyamatosan nő a berendezés- és helyigény.
- ◆ Egy cikk csak egy tárolóegységben érhető el. Ha több tárolóegységben helyezkednek el az együtt keresett kivágások, a kereső váltogatni kénytelen az egységeket.
- ◆ Egy egységhez, tehát 200 000 oldalnyi anyaghoz egyszerre csak egy munkatárs fér hozzá.
- ◆ Túl lassú az elérés. Egy cikk első oldala 15, minden további oldala 2–3 másodperc alatt érhető el. Gyakorlatilag minden cikket külön be kell állítani a képernyőn.
- ◆ Lassú a másolatkészítés: egy oldal 7 másodperc, de a xeroxozást többnyire meg kell ismételni, mert az expozíciót minden dokumentumra külön be kell állítani.
- ◆ Nem megfelelő a másolatok minősége. Gyakran lehetetlen teljesen olvasható másolatokat készíteni.
- ◆ Gyakori a tárolóegységek mechanikai meghibásodása. Állandó szervizre van szükség.

A másolatkészítés így nagyon leköti a személyzetet. Maga a keresés az adatbázisban csak perceket vesz igénybe. Ezután azonban ugyanazt a magasan kvalifikált munkatársat, aki a keresést végezte, lefoglalja az unalmas xeroxozás. Egy átlagos keresés 25–30 cikket, mintegy 40 oldalt eredményez. Ennek során csak a xeroxozás mintegy 20 percig tart, ami megengedhetetlenül hosszú idő a szerkesztőknek, akik *azonnal* meg akarják kapni az anyagot.

A mikrofilmlap-tároló egységek gyakori meghibásodása és a másolatok olvashatatlansága miatt a G+J továbbra is kénytelen fenntartani hagyomá-

nyos archívumát is. Ez azonban sok problémát vet fel. Egyre több időt vesz igénybe a besorolás és a rendben tartás. Egyre nő a helyigény, a dokumentációs osztálynak azonban nincs módja több tárolóhelyet létesíteni.

### A megoldás keresése

A dokumentációs osztály évekig kereste a gondokra a gyógyírt. Hamar világossá vált, hogy csak elektronikus megoldás jöhet szóba. A létrehozandó rendszernek a következő hatékonysági tényezőkkel kell rendelkeznie:

- ◆ Legalább 4 millió oldalnyi tárolókapacitás.
- ◆ A teljes anyag elérhető legyen a maximum 20 munkahely mindegyikéről.
- ◆ Képernyős megjelenítés.
- ◆ Gyors elérés.
- ◆ Gyors, kötegelt üzemmódú másolatkészítés, közvetlen emberi beavatkozás nélkül.
- ◆ Az adatátvitel lehetősége az eredeti dokumentumokra házon belül és házon kívül.
- ◆ Nagy üzembiztonság átlagosan napi 16 óras üzemidő mellett.
- ◆ A dokumentációs osztály korábbi szervezeti rendjébe problémamentesen lehessen beilleszteni.
- ◆ Kulcsrakész szállítás, mivel munkatársaink közül senki sem ért az elektronikus adatfeldolgozáshoz.

A Philips 1980-ban mutatta be a MEGADOC\* rendszert. Mivel ez a technika számunkra nagyon ígéretesnek tűnt, és a Philips is érdeklődést mutatott eredetileg irodaautomatizálási eszköznek szánt új terméke ilyen alkalmazása iránt, közös munkacsoportot alakítottunk a rendszerterv kidolgozására. 1983-ban szerződést kötöttünk a Philipszel egy kísérleti rendszer felállítására, hogy megvizsgáljuk, a MEGADOC rendszer kielégíti-e a G+J sajtóadatbank támasztotta igényeket.

### A kísérleti rendszer

A rendszert három lépésben hoztuk létre. Először 1984 márciusában egy minimális konfigurációt állítottunk fel, hogy azon beletanuljunk a rendszerbe és kipróbáljuk.

\* A Philips MEGADOC dokumentáló rendszere az optikai lemez három lehetséges változata, a csak olvasható, az egyszer írható és újraírható közül a másikon, az úgynevezett WORM- (egyszer írható, sokszor olvasható) változaton alapszik. — A ref.

1984 augusztusában állítottuk fel a beviteli rendszert. Ez lényegesen nagyobb, mint amekkora a folyamatos adatbevételhez kell majd. Úgy döntöttünk ugyanis, hogy 1972-ig visszamenőleg a teljes archívumunkat bevisszük a MEGADOC rendszerbe. Így az egész anyag egységesen lesz elérhető, nem lesz szükség az új rendszer és a hagyományos rendszer tartós párhuzamos üzemeltetésére.

Harmadik lépésként 1985 májusában állítottuk üzembe a keresőrendszert. Ennek 1986-ig tart a tesztelése. Ha ez sikeres, az új rendszert használatba vesszük, a hagyományos archívumot és a mikrofilmlaptárat felszámoljuk.

### Az elektronikus dokumentumtároló rendszer

A rendszerbe beviendő dokumentumok mindekelelt nyolcpozíciós, háromrészes azonosítót kapnak. Az első két pozíció a témakört jelzi (például P1 az NSZK politikája, K1 a kultúra). A következő négy pozíció (A000-tól Z999-ig) a kötegszám, az utolsó kettő (A0-tól Z9-ig) a dokumentumszám. Ugyanezt az azonosítót a keresőrendszerbe is bevisszük.

A képfelbontás a Telefax (fakszimile) rendszerre emlékeztet. A letapogató 4 millió képpontra bont fel egy oldalt. Két fényességértéket használ: egy képpont vagy fekete, vagy fehér. A kettőt egymástól elhatároló szűrkeérték állítható. Az így digitalizált információ nagyfelbontású képernyőn jelenik meg. Ha a kép minősége megfelel, hozzárendelődik az azonosító, és a rendszer mágneslemezen tárolja az oldalt.

A mágneslemezeiről a dokumentumokat kötegelten viszi át a rendszer az optikai lemezre. Az így felírt lemez a háttérpéldány, erről újabb másolással kerül az információ a tényleges tárolólemezre. A dokumentumok témájuk szerint más-más tárolólemezre kerülnek. Külön lemeze van a politikának, a gazdasági kérdéseknek, a kultúrának stb. Ezek a tárolólemezek "wurlitzer"-be kerülnek: olyan lemeztároló berendezésbe, amely automatikusan helyezi olvasópozícióba a kívánt lemezt.

A visszakeresés az adatbázisban való szokásos kereséssel kezdődik. Ez az IBM 3081-es nagyszámítógépen futó PDS rendszerben folyik. A talált dokumentumok kivonata képernyőn jelenik meg. Ha a kereső a dokumentumokat relevánsnak találja, *open document* parancsot ad ki. Ennek hatására a dokumentumok azonosítóját az IBM gép átadja a Philips számítógépnek. A wurlitzer olvasópozícióba helyezi az első szükséges optikai lemezt, a Philips rendszer kiválasztja az első dokumentum első oldalát és nagyfelbontású képernyőn megjeleníti. Ettől kezdve a

kereső teljesen az IBM gép termináljáról kezelheti a Philips rendszert: előre-hátra lapozhat a dokumentumban, átugorhat oldalakat, külön oldalakra vagy teljes dokumentumokra adhat ki nyomtatóparancsot. Amíg ő az első dokumentumot ellenőrzi, addig a Philips rendszer megkezdi az összes kiválasztott dokumentum átvitelét az optikai lemezekről mágneslemezre. Így bizonyos idő múlva a dokumentumokban a lapozás már nagyon gyors, a wurlitzer és az optikai lemezek olvasója pedig felszabadul egy másik kereső munkahely számára.

A keresés befejezése után a nyomtatás már automatikusan folyik.

## Tapasztalatok

Átfogó tapasztalataink pillanatnyilag még csak a beviteli alrendszerrel vannak.

1984 augusztusában két letapogatóból és egy ellenőrző munkahelyből álló konfigurációt állítottunk üzembe. Ezt 1985 elején egy harmadik letapogatóval egészítettük ki. Az első hetekben néhány hardver- és szoftverhiba hátráltatta a munkát. 1984 októberére vittük fel az első százezer oldalt, 1985 májusára már eljutottunk a félmillió oldalig.

Egy oldal letapogatása 4,5 másodpercig tart, de ha valamennyi letapogató egyidejűleg dolgozik, akkor ez az idő 6 másodpercre nő. Ebben benne van az oldal fölhelyezése, a tényleges letapogatás, a digitalizálás, az adatkompresszió, a mágneslemezre írás, a visszaolvasás, a dekompresszió, végül a nagyfelbontású képernyőn való megjelenítés.

Az archív anyag felvitele ennél lassúbb, mert sorrendi hibákat kell kiigazítani, témakörökbe kell osztályozni a dokumentumokat, majd vissza kell sorolni őket a hagyományos archívumba. Ebből az anyagból napi 3500–4000 oldalt tudunk feldolgozni.

Eleinte az egyik egységen felírt optikai lemezt nem tudtuk elolvasni a másik egységen. Ez azonban csak beállítási hibából eredt, amit sikerült kijavítani. Az optikai lemezekkel komolyabb bajunk csak egyszer volt, amikor egy szállítmány hibás lemezt kaptunk. A letapogatás után 10:1 arányú adatkompressziót tudtunk elérni, így 50–60 kb-át szükséges egy oldal tárolásához, mintegy 40 000 oldal fér egy kétoldalas optikai lemezre.

A letapogatott dokumentumokkal nincs sok baj. Mintegy 5%-ban kell megváltoztatott beállítással megismételni a letapogatást. Csak néhány dokumentum volt, amelyről végképp sikertelen volt a letapogatás. Ezek többnyire színesek voltak, például zöld papíron kék betűk. A sikertelen felvételek aránya lényegesen kisebb, mint a mikrofilmzésnél.

A nagyfelbontású képernyőn való megjelenítés minősége kitűnő, kontrasztos, az apró betűk is jól olvashatók, a diagramok jól kivehetőek.

A képernyő\* is kitűnő. Percenként 20 oldalt nyomtat, halk, megbízható. Jelenleg a négy kereső munkahelyhez két nyomtató tartozik.

## Következtetések és előtekintés

1986 első néhány hónapjában végezzük el a keresőrendszer tesztelését 1 000 000 oldalnyi anyagon. Ennek során különös figyelmet fordítunk a válaszidő és a megbízhatóság vizsgálatára.

Ha a tesztelés pozitív eredménnyel zárul, akkor a kereső alrendszert 12 munkahelyesre bővítjük. A továbbiakban közvetlen Telefax- (faksimile-) csatlakozást létesítünk, hogy a dokumentumokat telefonvonalon továbbíthassuk a felhasználóknak. A MEGADOC ehhez a CCITT 3 Telefax-szabványt alkalmazza. 1987-re a teljes archív anyagot bevisszük a rendszerbe, 1972-ig visszamenőleg.

Egy kísérleti program mindig időigényesebb, mint egy kipróbált rendszer installálása. Mi is egy évet késtünk eredeti tervünkhöz képest. Cserébe viszont igényeinknek megfelelően sokkal jobban tudtuk befolyásolni a konfiguráció kialakítását. Végül is az első keresési tapasztalatok szerint megérte.

## Irodalom

- KLIMBIE, J. W.: Digital optical recording, principles and possible applications. = 5th International Online Information Meeting, London, December 1981, Learned Information, London and New Jersey, 1981, p. 315.  
SCHMITZ-ESSER, W.: Digital homograms to offset automated microfiche access for a very large databank – an outline of a pilot project by Gruner und Jahr Publishing based on Philips DOR technology. = *ibid.* p. 323.

/HARTMANN, K.–LIENAU, H.-J.: Gruner und Jahr press data bank – first experiences of usage of optical disc storage technology (MEGADOC) for a large archive. = 9th International Online Information Meeting, London, 3–5 December 1985. Learned Information, Oxford and New Jersey, 1985. p. 237–241./

(Válas György)

\* Valószínűleg lézernyomtató. – A ref.