

A Fórum Intézet és a digitalizálás

Gyakorlatilag az ember egész környezetében digitális adatok, digitálisan vezérelt gépek, digitális adathordozók egyszerűsítik – vagy bonyolítják – a mindennapi munkát. Tekintve működési elvük egyszerűségét, előállításuk olcsóságát és a megtakarított idő felbecsülhetetlen mennyiségét, a digitális berendezések az élet majdnem minden területéről kiszorították az analóg, illetve a mechanikus gépeket. Arra való tekintettel, hogy a digitális gépekben nincs szükség mozgó alkatrészekre (ezáltal megszűnnek a kopás okozta pontatlanságok), a gépek teljes élettartamuk alatt ugyanarra a minőségű munkára alkalmasak. Ez a munkafolyamat a végeredményénél tűnik felbecsülhetetlen értékűnek.

Bár a digitális gépek élettartama (fejlődésük nagy ütemének köszönhetően) jelentősen kevesebb, mint analóg elődeiké (annak idején egy kazettás magnetofont még tíz-tizenöt év után is lelkesen használtunk, bár a kazetták már agyon voltak gyűrve, a rajtuk lévő hanganyag már-már élvezhetetlenné vált), az újabb, jobb minőségű gép megvásárlása anyagilag kevésbé megterhelő, sőt sok esetben nem is szükség szerű. Továbbá lehetőségünk nyílt a gépek teljesítményének, munkájuk minőségének teljesen az igényeinkhez való igazítására, a költségek emelkedése nélkül, ami természetesen elődeiknél szinte lehetetlen volt.

A számítógépek ma is használt felépítése egy magyar származású matematikus, Neumann János munkája. A budapesti születésű matematikus matematikai tehetsége már a pesti evangélikus gimnáziumban kibontakozott, édesapja mégis azt kívánta, hogy vegyészmérnök legyen belőle. Ezért az érettségi után beiratkozott a zürichi műegyetemre, miközben Berlinben és Budapesten matematikát, fizikát és filozófiát hallgatott. Csaknem egyidőben szerezte meg svájci mérnöki diplomáját és a budapesti matematikai doktorátust. Neumann a mérnöktovábbképző intézetben (Institute for Advanced Studies – IAS) lett a matematika professzora. A második világháború alatt Neumann jelentkezett a Manhattan-terv végrehajtására, az amerikai atombomba előállítására; amelynek kapcsán a lökéshullámokkal kapcsolatos számításokat végezte. Ebben az időben már elkészült néhány jelfogós számítógép, mint például a MARK-I, valamint rendelkezésre állt egy mechanikus analóg számítógép, a Bush-féle differenciálanalizátor, és számos – akkor általánosan használt – elektromechanikus számológép is. Sajnos, egyikkel sem lehetett másodpercenként néhány műveletnél nagyobb sebességet elérni.

Neumann 1944-ben ismerkedett meg Hermann Goldstine-nal, aki az aberdeeni kísérleti lőtérén a lö- és bombázási táblázatok számításait irányította. Tőle megtud-

ta, hogy a hadsereg a Pennsylvanai Egyetemen egy elektronikus számolóberendezés – az ENIAC – kifejlesztésén dolgozik, s azt 1000 művelet/másodperc sebességre tervezik. Neumann-nak épp egy ilyen gépre volt szüksége. Egy évvel később a projekt igazgatója lett. Neumann és Goldstine ettől kezdve – egészen Neumann haláláig – együtt dolgozott a számítógépek fejlesztésén. Az ENIAC két tervezőjével, John W. Mauchlyval és Prespert J. Eckerttel azonban együttműködésük nem volt túlságosan szoros. Neumann nem járult hozzá ahhoz, hogy az ő általa feltalált tárolt program elvét a négyük nevén szabadalmaztassák. Hogy megakadályozza Mauchly és Eckert szabadalmaztatási szándékát, közölte a leírást, s ezzel lehetetlenné tette a szabadalmaztatást.

Neumann és Goldstine Princetonban megalkotta az IAS számítógépet. Ez leginkább abban különbözött a korábbi két számítógéptől, hogy párhuzamos működésű volt, tehát sokkal gyorsabban számolt bármelyik korabeli számítógépnél, felépítése pedig – fő vonalaiban – megegyezett a mai modern számítógépekével. Neumann váratlanul betegedett meg, kiderült, tüdőrákja van. Betegsége alatt is folyamatosan *A számítógép és az agy* című előadásán dolgozott. A kézirat, sajnos, befejezetlen maradt, Neumann 1957. február 8-án meghalt. Életét és munkásságát mára szinte teljesen a számítógéppel kötik össze, pedig számtalan matematikai definíció megfogalmazása kötődik még a nevéhez.

Neumann munkásságának köszönhetően vált lehetővé az adatok óriási mennyiségű tárházának létrehozása, hihetetlenül kis helyen. Amikor adatok tárolásáról beszélünk, alapvetően logikai értékek tárolásáról van szó. Érvényes ez a szövegek, a képek, a hangfelvételek, a filmek esetében is. Míg egy képnél az egymást követő különböző színű képpontokat kell megkülönböztetni, ahol egy pont értéke a színmélységének felel meg, addig a hangoknál a pont értéke frekvenciát jelöl. A kép megjelenítéséhez a grafikus kártya dekódolja az állomány tartalmát, majd azt a képernyőre vetíti, a hanganyagnál a hangkártya végzi a dekódolást, a végeredményt pedig a reproduktorokból hallhatjuk. A film nem más, mint a szem érzékenységénél nagyobb sebességben (ami kb. 50 Hz) egymást követő képek halmaza, amihez azonos időtengelyen egy hangállományt csatolunk. A szöveg előre definált karakterek (ASCII) dekódolt megjelenítése.

Mivel a számítógépek és a rajtuk használt operációs rendszerek nem egységesek, sőt a kompatibilitás is csak nagyvonalakban szabványosított (például egy Linux alatt tömörített állományt nem triviális Windows alatt kicsomagolni), a digitalizált állományokat ajánlatos olyan állományokra alakítani, amelyeket bármilyen operációs rendszer kezelni tud.

A Fórum Intézet és a digitalizáció

Intézetünk öt éve kezdett az adatok digitalizációjával foglalkozni. Ekkorra vált lehetővé olyan számítógépek, illetve perifériák beszerzése, melyek megfelelő minőségben képesek az adatok feldolgozására. Természetesen a technika folyamatos fejlődésének köszönhetően mára ezek a berendezések elavultak, helyükbe újabb, jobb minőségű berendezések kerültek.

Ma az adatok feldolgozását egy helyen, az intézetben kialakított irodában végezzük. Három számítógép, két hordozható számítógép alkotja a digitalizáláshoz szük-

séges berendezések alapját. Ezek mindegyike alkalmas képek feldolgozására, szkennelésére. Két számítógépet egy 1,1 GHz-es processzor hajt, memóriájuk 256 MB SDRam, 133 MHz-es sínen. Mindkettőhöz egy negatívok szkennelésére is alkalmas szkennert van csatlakoztatva. Ezeket a gépeket használjuk képek bevitelére. A feldolgozásuk már rendszerint a hordozható számítógépeken történik, ezzel optimalizálva a gépek munkaidejét. Harmadik számítógépünk gyorsabb, 2,1 GHz-es Celeron-processzorral van felszerelve, belső sínjeinek frekvenciája 266 MHz, memóriája 512 MB DDR. Ebbe a számítógépbe egy Creative hangkártyát helyeztünk, amely remek paramétereivel alkalmas a hanganyagok megfelelő minőségben történő feldolgozására. A kommunikációt a gépek között 100 Mbit-es hálózat oldja meg. Jelenleg egy drótnélküli hálózat kiépítésén dolgozunk. Bár ennek a sebessége jelenleg kisebb, mint 100 Mbit, a jövőben valószínűleg ez változni fog, kiépítése és karbantartása pedig költségkímélőbb, valamint lehetővé teszi az internet használatát bárhol az épületen belül.

Az adataink interneten történő elhelyezését egy saját szerver üzemeltetésével oldottuk meg. Ez a szerver egy internetszolgáltatónál van elhelyezve, paramétereit tekintve egy kétprocesszoros Xeon típusú számítógép, 2 GB DDR memóriával, 533 MHz-es sínnel, két, párhuzamosan használt merevlemezzel, ami az adatvesztés ellen nyújt biztonságot.

Mára sikerült elérnünk, hogy a most készülő felvételek digitális rögzítőkkel készülnek, így azok feldolgozása minimális időt vesz igénybe. A digitális fényképezőgép alkalmas a konferenciák, a képzések dokumentációjának elkészítésére, a digitális hangrögzítők pedig a beszélgetések, az előadások hanganyagának rögzítését teszik lehetővé.

Képek feldolgozása

A képek archivációja, digitalizációja több, párhuzamosan futó munka eredménye. Először a képek nagy felbontású szkennelése történik, későbbi akár nyomdai felhasználásra is alkalmas méretben. Ez mérettől függően 300-600 DPI (Dot per inch – képpont per hüvelyk), 16 millió színmélység. Negatívok szkennelésénél ez a felbontás még nagyobb, 2400 DPI. Sok esetben már az eredeti kép minősége sem kielégítő, ezért a képek feldolgozására speciális képszerkesztő programokat használunk. Ezek túlnyomó része ingyenesen letölthető az internetről. Lehetőséget kínálnak többek között a kép élességének javítására, a forgatásra, a színek élesítésére, a képméret változtatására. Egy sérült, alig kivehető negatív esetén kisebbfajta csoda végezhető velük.

Javítás után a képek egy speciális kódot kapnak, amely alapján visszakereshető az eredeti példány, ha arra szükség van. Az interneten való eléréshez ezekből a képekből két, minőségében és méretében kisebb kép készül, informatív jelleggel. Ezek méretre 150, illetve 500 képpont szélességűek, és arányaiban hozzáigazodó magasságúak. Felbontásuk 72 DPI. Nyomdai felhasználásra alkalmatlanok, viszont nagy sebességű keresést, illetve megjelenítést tesznek lehetővé a világhálón. A képek eredetije, amennyiben nem jogvédett, bárki számára postázható vagy személyesen megtekinthető a Bibliotheca Hungarica archívumában.

Az ezt követő munka a képekből készített metaadat, tartalomleíró állomány. Ezek készítésével intézetünk mégcsak nemrégiben ismerkedett meg, digitalizált képeink nagy részéhez mégis sikerült elkészítenünk ezeket a leíró állományokat. A leíró állomány lehetővé teszi a platformfüggetlen (értsd operációs rendszertől független) keresést, az állomány tartalmának megtekintését. Alapul a Dublin Core-szabványt használjuk, melyet XML-állományba ültetve készítjük az egyes képek leíró állományait. A kereséshez az NDA (Nemzeti Digitális Adattár) kereső motorját használjuk saját weboldalunkon is. Az XML-állományokat rendszeresen továbbítjuk az NDA adatbázisába, így adataink itt is megtekinthetők.

Jelenleg a feldolgozott, leíró állománnyal ellátott képek száma:

- Fórum Intézet (konferenciák, klubok stb.) – 7287 db
- Bibliotheca Hungarica képeslapgyűjteménye – 2386 db
- Szakrális Kisemlék Archívum – 281 db
- Csemadok-archívum (digitalizálva) – 7681 db
- Gyökeres György képarchívuma – 452 db
- Arany A. László zoboralji képeinek gyűjteménye (színes dia) – 680 db

Feldolgozásra vár:

- Csemadok-archívum
- Szóttas-archívum
- Egyéni hagyatékok
- Arany A. László zoboralji képei (újabb 46 csomag, kb. 2500 képkocka)

A Fórum Kisebbségkutató Intézet fotóarchívumát elsősorban az intézet életével kapcsolatos felvételek képezik, a különböző konferenciák fotódokumentációja, a klubesteken készített felvételek és egyéb, az intézet életével kapcsolatos felvételek.

A Bibliotheca Hungarica képeslapgyűjteménye teljes egészében feldolgozásra került, megtekinthető az interneten. Ez a gyűjtemény elsősorban a dél-szlovákiai falvak, városok által megjelentetett képeslapokat foglalja magába. Állandóan bővül, az új darabok szinte azonnal felkerülnek az internetre. A világhálón 500 képpont szélességű méretben található meg, a kiválasztott darabból másolatot a Bibliotheca Hungaricában lehet beszerezni. A gyűjtemény egy része nálunk csak digitális formában található meg.

Az Etnológiai Központ már évek óta foglalkozik a dél-szlovákiai szakrális kisemlékek fényképezésével. Gyűjteményük egyedülálló. Az utóbbi években már digitális fényképezőgép segítségével örökítik meg az utak mellett található kereszteteket, a szentek szobrait. A régebben készült felvételek folyamatosan kerülnek feldolgozásra. Munkájuk eredménye több kötetben is megjelent.

Gyökeres György hosszú évekig fotózással foglalkozott. Az ő személyes archívuma gyűjteményünk egy értékes darabja. Felvételein elsősorban a Csemadok által szervezett fesztiválok, programok jelennek meg, de az ő érdeme a felvidéki írók portrégyűjteménye is. Ez a gyűjtemény teljes egészében digitalizált, TIFF-formátumban, nagy, 600 DPI-s felbontásban az intézet archívumában található meg.

Arany A. László a második világháború éveiben foglalkozott a zoboralji falvak fotózásával. A felbecsülhetetlen értékű képanyag negatívok formájában áll intézetünk rendelkezésére. A felvételeken megtalálható az aprólékosan lefotózott női, illetve férfiviselet, több zoboralji faluból, valamint a korabeli élet apró mozzanatai, a min-

dennapi munka képei. A gyűjtemény a zoboralji élet egy értékes archívuma. Az anyag nagy része már feldolgozásra került, több kiállítás alapját is képezte, elsősorban a Zoboralján. Nemrégiben újabb, több mint 2000 képet számláló csomag került intézetünkbe, ezek digitalizálása folyamatban van. A teljes képanyag részletes, átfogó gyűjtemény a zoboralji falvak életéről.

Hanganyagok feldolgozása

A hanganyagokat a már fent említett Creative típusú hangkártya segítségével dolgozzuk fel. Az előnye abban rejlik, hogy képes akár 96 kHz-es tartományokat is rögzíteni, ellenben a klasszikus 44 kHz-es hangtartománnyal. Ez jelentős minőségbeli javulást eredményez. Bár az emberi fül számára a határt a 20 kHz-es hangtartomány képezi, mégis valamiként érzékeljük az e fölötti tartományokat is (gondolok itt elsősorban a bakelitlemezek sajátos sűgására, ami digitalizációkor más típusú kártyával elveszik). Természetesen a hangkártya e tulajdonsága csak indokolt esetben van kihasználva, tekintve, hogy a hangállományokat sűríteni kell, ezáltal nyújtva lehetőséget az internetes elérésre. Sűritésre jelenleg az MP3-as formátumot használjuk, 192 kbit/s-os adatsűrés mellett.

Az archívum részét képezik:

- Sima Ferenc gyűjtései (nyelvjárási szövegek);
- Csemadok-gyűjtések (nyelvjárási szövegek);
- a Pátria Rádió által készített riportok;
- egyéni hagyatékok, felvételek;
- Oral History (beszélgetések Szlovákiában élő írókkal, politikusokkal a rendszerváltás éveiről).

Kihasználva a digitális hangrögzítők előnyeit, az Oral History-sorozat már digitálisan is készül. Az ezekhez tartozó leírások párhuzamosan a felvétellel készülnek. Tekintve eme riportok méretét, egyelőre a világhálón nem elérhetők, ez a jövőbeli feladataink része. A tárolás MP3-as formátumban, 44 kHz-en történik. Itt nincs semmiféle minőségbeli romlás az analóg felvételhez képest, tekintve, hogy az ember hangtartománya 2000–5000 Hz közé esik.

Sima Ferenc 118 szalagból álló nyelvjárási gyűjteménye csaknem teljes egészében digitalizálásra került, az egyes szalagok több külön állományra lettek felosztva. Ezek mérete és időtartama a riportok hosszának felel meg. Egy szalagon átlagosan 1,5-2 órányi hanganyag lett rögzítve. Némelyikük egységes elbeszélés, a többin különböző riportalanyok szólnak meg egy adott községből.

A Csemadok Népi Köre a hatvanas évek végétől a nyolcvanas évek elejéig készített felvételeket, falusi közegben. Az elsősorban elbeszéléseket, az élet mindennapjait tartalmazó felvételeket 61 magnószalagon rögzítették. Sajnos pontos leírás ezekről a hangállományokról nem áll rendelkezésünkre, így csak azt tudtuk leírni, ami a magnószalagokon elhangzott. A gyűjtemény nem teljes, 47 szalag került digitalizálásra.

A Pátria Rádió által felvett riportok 193 magnószalagon találhatók, egyelőre feldolgozás alatt. A vágott, sugárzott riportok hossza 10-15 perc, de sok esetben a vágatlan anyag is megtalálható archívumunkban. Ezekről a felvételekről – hála a Pát-

ria Rádió munkatársai precizitásának – pontos adataink vannak, feldolgozásuk az elkövetkező hónapokban megtörténik.

Szerverünk kapacitása egyelőre nem teszi lehetővé a hanganyagok elérését a világhálón. A teljes anyag mérete óriási, több száz GB. Ezek szerveren való elhelyezését a jövő évben tervezzük.

Kiadványok

A könyvkiadás, a publikációk megjelentetése a Fórum Kisebbségkutató Intézet munkájának egy fontos része. Könyveink több sorozatban jelennek meg, illetve 1999-től rendszeresen megjelenik a *Fórum Társadalomtudományi Szemle*, amelyben tudományos jellegű írások kapnak helyet. A Magyar Elektronikus Könyvtárnak (MEK) PDF-formátumban küldjük kiadványainkat, így azok az ő weboldalukon is olvashatók.

Az Arcanum-sorozat példája alapján mi is szeretnénk saját kiadványainkat digitális formában is megjelentetni. Ez egy szövegkeresővel ellátott adatbázis kialakítását igényli. A könyveket és tanulmányokat ebben az adatbázisban elhelyezve, sorozatonként tennénk közzé, ezáltal megkönnyítve a keresést és a tájékozódást az egyes szövegekben. A program kialakítása folyamatban van, jelenleg az egyes könyvek feltöltésével foglalkozunk.

Konferenciák

A Fórum Intézet rendszeresen rendez konferenciákat, értelmiségi fórumokat, klubesteket. Ezek rögzítése, dokumentálása szintén részét képezi a munkánknak. Az itt elhangzott előadásokat a kezdetekben digitális magnó segítségével rögzítettük, majd később kerültek tényleges feldolgozásra, vágásra.

Mára kifejlesztettünk egy felvevőrendszert, amely lehetővé teszi a konferenciák anyagának reális időben történő feldolgozását. Ennek segítségével a konferencia végeztével elkészül a konferencia teljes képanyagát és az elhangzott előadásokat tartalmazó CD, amit bárki magával vihet, használhat.

Ez a berendezés gyakorlatilag három számítógép, egy digitális fényképezőgép és a meghívó alapján előre elkészített programfelület. Míg az egyik számítógép a hanganyagot rögzíti, addig a másikon az előző előadás vágása, feldolgozása folyik, a harmadikon pedig a már elkészült anyag betöltése a programfelületbe. Ennek a módszernek a legnagyobb előnye, hogy az egyes konferenciák anyagai azonnal archiválódnak, kiküszöbölve ezzel az emberi feledékenység okozta adatvesztést. A későbbiekben ez az anyag is egy adatbázisba kerül, majd egy keresővel egészül ki. Adathordozón és interneten egyaránt elérhetővé válik.

Levéltár és könyvtár

Nemrégiben indult el egy folyamat, amely a nyomtatásban megjelent szövegek feldolgozását segítené. A Bibliotheca Hungarica könyv-, illetve folyóirat-gyűjteménye jelenleg nyomtatott formában áll a kutatók rendelkezésére. Ezek digitalizálása eddig gyakorlati okok miatt nem történt meg. A szövegek újragépelése lehetetlen vállalko-

zásnak tűnik, úgyhogy egyedül a szkennelés jelentene megoldást. Csakhogy így a szöveges dokumentumok képként őrződnek, lehetetlenné téve egyfajta szövegkéréső létrehozását. A szövegfelismerő szoftverek sajnos túl nagy hibaarányal dolgoztak ahhoz, hogy egy ekkora adatmennyiség mellett érdemben használni lehessen őket. Erre a problémára a Neumann Klub informatikai köre ajánlott nekünk egy lehetséges megoldást. A Google ingyenes szövegfelismerő motorját alakítják át a Bibliotheca Hungarica igényeinek megfelelően. A gyakorlatban ez azt jelenti, hogy minden nyomtatásban megjelent folyóiratra egy külön szoftver lesz kialakítva. Erre azért van szükség, mert a különböző folyóiratok különböző elrendezésűek, valamint különböző nyomdákban készültek, ezáltal különböző betűtípusokat használtak. Ezzel a módszerrel előzetes számításaink szerint akár 98%-os pontosságú szövegfelismerést érhetnénk el, ami hihetetlen módon gyorsítani tudná a digitalizálás folyamatát. Ehhez még egy A/3-as méretű szkennerre is szükség van. Az újabb, jobb állapotban lévő dokumentumok, folyóiratok akár nagy teljesítményű lapolvasóval is feldolgozhatók. Egynémely dokumentum mostanra sajnos olyannyira megrongálódott, hogy ezek bevitele csak külön-külön, laponként elképzelhető. Végeredményben a nem jogvédett szövegek az internetre kerülhetnének, a regionális vagy régebbi folyóiratok gyűjteményként DVD-n is elérhetővé válnának bárki számára.