

LUX ZOLTÁN

A DIGITÁLIS TÖRTÉNETI ADATARCHÍVUM¹

Manapság egyre inkább terjed az informatikai eszközök használata a tudományos világban, az oktatásban és az ismeretterjesztésben. És ez – úgy tűnik – kezdi alapvetően átalakítani a hagyományos struktúrákat. Egyre közelebb hozza az alap kutatásokat az alkalmazásfejlesztéshez, megszünteti a „tisza tudományok” zártságát. Lerövidíti a kutatási eredmények és azok alkalmazása között eltelt időt, közelebb hozza egymáshoz az egyes szaktudományokat. Az eredmények alkalmazhatósága érdekében már nem a téma tisztán diszciplináris, hanem inkább interdiszciplináris megragadása a cél.

Az oktatás és ismeretterjesztés terén is hasonlóan lényeges változások mennek végbe. Az iskola többé nem a tudás megszerzésének egyedüli helye, s a tanár nem annak egyedüli közvetítője. Az internet, a digitális adathordozók által közvetített tudás megteremtheti, meg kell teremtenie a művelődés, tudás karbantartás megfelelő virtuális környezetét. Ennek a környezetnek lehetnek részei, építőelemei azok a digitális adatarchívumok, amelyek képesek a felhalmozott tudást egy folyamatosan változó környezetben megőrizni, a változó alkalmazási igényeknek megfelelően átszervezni.²

AZ ADATBÁZIS-ÉPÍTÉS CÉLJA

Miért építünk adatbázisokat mi is, mire jó ez a sokszor kicsit nehézkesnek tűnő és költséges technológia a történettudományban? Az ma már nyilvánvaló és természetes, hogy a számítógépek segítségünkre vannak tanulmányaink elkészítésében és formába öntésében. Hasznos lehet a világhálón megnézni, hogy melyik könyvtárban találjuk meg a szükséges szakirodalmat. A programok megkönnyítik a szerkesztést, a javítást, s ha kész a cikk, elektronikus levélben küldjük el valamelyik szerkesztőségbe. De miért kell a kutató oldaláról is betagolódni az adatokat gyűjtő és azokat valamilyen módon másoknak is hamar elérhetővé tenni akaró rendszerbe? Néha úgy érezhetjük, hogy a technológia hajt minket, kényszerít valamire, aminek a szükségességéről nem is vagyunk meggyőződve.

1 Ez a munka az Improving Human Potential (IHP) and the Socio-economic Knowledge Base – Enhancing Access to Research Infrastructures támogatásával létrejött, a kvalitatív adatarchiválás informatikai problematikáival foglalkozó tanulmány első része. A kutatási ösztöndíjat az ECASS (European Centre for Analysis in the Social Sciences) biztosította 2001 és 2003 között.

2 Az informatika fejlődésének a kutatásra, művelődésre gyakorolt hatásával rendkívül részletesen foglalkoznak Nyíri J. Kristóf tanulmányai: <http://www.phil-inst.hu/nyiri/>

Ha a folyamat távolabbi jövőbe mutató kérdését nem is tudjuk megválaszolni, néhány érvet azért felvázolhatunk az adatbázis-építés mellett.

Az összegyűjtött forrásokat, adatokat használni akarjuk kiadványaink összeállításában. Könnyű belátni, hogy ha például csak illusztrációként van szükség egy-egy fotódokumentumra, már akkor is érdemes a meglévő technológiát használva azt egyszer és mindenkorra tárolni, és visszakereshetővé tenni egy későbbi esetre vagy kollégáink számára.

Különösen fontos a rendszerezett tárolás multimédia-kiadványok, interneten elérhető publikációk készítése esetén. A digitális kiadványokban lényegesen több forrást, szemléltető anyagot, multimédia-dokumentumot lehet felhasználni, mint a nyomtatottban, s ebben az esetben a visszakereshetőség és a digitális forma is fontos szempont.

Kronológiát természetesen lehet szerkeszteni, karbantartani kartonokkal teli dobozok segítségével is. Azonban ha nem csak időrendben szeretnénk keresni a „dobozban”, akkor anynyi dobozra van szükségünk, ahány szempont alapján lehetővé kívánjuk tenni a keresést. Szükségünk lehet például azokra az eseményekre, amelyekben X. Y. részt vett, vagy azokra, amelyek Z helyen történtek, vagy például szeretnénk kigyűjteni azokat az eseményeket, amelyekre vonatkozóan W forrásból merítettünk adatot. Jól illusztrálják a probléma lényegét a könyvtári katalógusok. Kezelésüket a számítógépes keresőrendszerek már jó ideje megkönnyítik.

Az adatbázis-építés célja tehát, hogy a forrásokat könnyebben használhatóvá, újrafelhasználhatóvá tegye, és a publikálás számára segédanyagot, szemléltető dokumentumokat biztosítson.

TÁVLATI CÉLOK

Szó esett már események, források, képi dokumentumok összekapcsolásáról. Azaz az adatbázisban „azt mondjuk” például, hogy kapcsoljuk össze Kádár János életrajzát egy csomó képpel, amelyet különböző archívumokban róla találtunk, majd feldolgoztunk, kapcsoljuk össze azokkal az eseményekkel, amelyekben részt vett, és azokkal a forrásokkal, szakirodalmi publikációkkal, interjúkkal, amelyek vele kapcsolatosak. Ha ebben a hálóban még történeti (társadalmi, politikai stb.) narrációk is vannak, az előálló adatbázis lényegében egy folyamatosan fejlesztett multimédia-történelemlépcső, amelynek a fejlesztését és hozzáférését informatikai eszközökkel több szinten is lehet szabályozni. Ha a történelemlépcső-adatbázisnak vannak még nem teljesen kész részei, azokat nyilván nem kell általánosan elérhetővé tenni: előválogatások végezhetők a különböző szintű hozzáférések (ismeretterjesztés, közoktatás, felső szintű oktatás, kutatás) szempontjai szerint. A háló elemeit el lehet látni például arra vonatkozó címkékkel, hogy az adott tételt (rekordot) csak a felső szintű oktatástól „felfelé” tartalmazza a „történelemlépcső”. Az adatbázis egyes elemei lehetnek „fizetősek”, például egy fotódokumentum nagyobb felbontású, publikálásra is alkalmas változatának letöltéséért térítést lehet kérni. Az adatbázis tartalmazhat személyiségi vagy szerzői jogi szempontból nem nyilvános, nem általánosan hozzáférhető elemeket is. Ezek a megszorítások informatikai eszközökkel tökéletesen kezelhetők.

A kutatók talán kissé idegenkedve gondolnak arra a lehetőségre, hogy kutatási részeredményeiket is elérhetővé tehetnék egymás számára. Ez sok tévedéstől óvhatná meg a történettudományt, hiszen nincs az a kutató, aki a hagyományos kutatások során valamennyi elérhető forrást átnézte volna akár csak egyetlen témában is. Természetesen érthető, ha a kutatók a fontos történeti kérdésekkel kapcsolatban felbukkanó, sokszor véletlenszerűen megtalált do-

kumentumok első publikálását nem szeretnék kiengedni a kezükből. A levéltárak állományainak digitalizálásával azonban – valamint annak eredményeként, hogy a vállalatoknál, a kormányzatnál egyre gyakoribb, hogy egyes dokumentumok már csak digitális formában jönnek létre – a források megtalálása már nagymértékben a rendelkezésre álló információtechnológiai eszközök értő és okos alkalmazásától függ. A forrás minden érdekelt számára – valószínűleg interneten keresztül is – elérhető lesz. Ilyen összefüggésben talán nem a megtalált források őrzése lesz a hasznos, hanem sokkal inkább a történeti problémákra való gyors reagálás. Nem kisebb, hanem éppen lényegesen nagyobb szerepet kaphat a történetész értelmezés, hiszen a kutatói aprómunkával nem kell annyit bíbelődni.

Az „okos szoftverek” esetünkben „adattányászati”, illetve tudásmenedzsment- (*Knowledge Management*) szoftvereket jelentenek. Ilyen szoftvereket az üzleti világban már évek óta alkalmaznak. Adattányászati szoftverekkel dolgoznak például a pénzügyi szférában a rendelkezésre álló adatállományból a felhasználó által előre nem definiált elemzések elkészítésére. Ez a használat természetesen a program lehetőségeinek mélyebb ismeretét feltételezi. Jóllehet a levéltárak zöme még állománya feltárásának a kezdetén áll, már van olyan levéltár, amelyben a rendelkezésre álló dokumentumok mély tartalmi feltárása folyik, beleértve a dokumentumok képi tárolását is.³ És ne feledjük, jelenkortörténetünkkel kapcsolatban egyre több digitális dokumentum keletkezik, gondoljunk például a már említetteken kívül még a törvényekre vagy a *Parlamenti Naplóra*.⁴

Az adattányászati szoftverek továbbfejlesztései az említett tudásmenedzsment-szoftverek. Ezen a területen nagy lendülettel folynak a kutatások és fejlesztések. A világ valamennyi fejlett országa – az Európai Unió is – kiemelten támogatja a tudásmenedzsment körébe tartozó kutatás-fejlesztési tevékenységeket. E kutatások során létrejövő termékek funkcióját a következőképpen lehet leírni: egy szervezet hatékony és eredményes működése alapvetően az ott felhalmozott tudás nagyságán és annak hatékony felhasználásán múlik, tehát a tudásmenedzsment-programoknak az a feladata, hogy összegyűjtse a munkatársaknál található információkat, és a szükségeseket automatikusan és a lehető leggyorsabban eljuttassa azokhoz, akiknek a munkáját segítheti.⁵

Mind az adattányászati, mind a tudásmenedzsment-eszközök fejlesztése a mesterséges-intelligencia-kutatások jelentős fejlődésén alapul. Ezekben az esetekben ugyanis már messze többről van szó, mint arról, hogy egy karakterláncot (keresőkifejezést) egy tetszőlegesen nagy adathalmazban megtaláljunk, azaz keresést hajtsunk végre. Itt már valamiképpen meg kell „érteni” a szöveget, „ki kell találni”, hogy a felhasználó mire gondolhatott. Anélkül, hogy a matematikai részletekbe belemennénk, egy a társadalomtudósok számára is könnyen érthető problémát szeretnék felvázolni. Sok társadalomtudós használ kutatásai során szövegelemző (*Text Analysis*) szoftvereket. Több ilyen szoftver van, és valamennyi specializálódott bizonyos feladatokra⁶. Egyben azonban megegyeznek, mégpedig abban, hogy lényegében „kézzel” és

3 A Történeti Hivatalban 2001-ben megkezdődött az ott őrzött állambiztonsági iratok nagyon részletes tartalmi feltárása, amelynek keretében a teljes dossziék tartalma is az adatbázisba kerül. Ezzel meggyorsul a dokumentumokhoz való hozzáférés technikai folyamata.

4 Az érvényes törvények gyűjteménye már több mint tíz éve kapható rendszeresen frissített formában CD-ROM-on, több kiadó fejlesztésében is. A törvények feldolgozása, digitalizálása visszamenőlegesen is folyik. Történeti szempontból legalább ennyire fontos a különböző parlamenti iratok digitális archiválása. A *Parlamenti Napló*-ban 2002 őszétől már a felszólalások digitalizált videorészletei is megtekinthetők.

5 A kérdéskörnek jó kifejtését adja a *Knowledge Management Handbook*. (Szerk. Jay Liebovitz. London–New York–Washington D.C., 1999, CRC Press, Boca Raton) 8. fejezete.

„egyenként” kell az egyes szövegrészekhez hozzárendelni azoknak az adott kutatás szempontjából releváns „jelentését”. Ha ez megvan, akkor a program nagyon gyorsan elvégzi az elemzést, a fogalmak szövegrészekhez rendelése azonban meglehetősen hosszadalmas munka lehet, amely bizony az elemezhető szövegek hosszát jelenleg még erősen behatárolja. A kutató álma az lenne, hogy ő valamilyen paraméterrendszerrel megadja az elemzés szempontjait, azután a program értelmezi a szöveget, mintegy az egész szöveget „elolvasva” megtanulja az adott szöveg sajátos kifejezőrendszerét, és azt hozzárendeli a megadott szempontrendszerben lévő fogalomrendszerhez. Ezt természetesen megteheti több lépcsőben, mintegy interaktív kapcsolatban a kutatóval. Ilyen program ma még nem kapható, tudományos konferenciákon azonban már beszámolnak hasonló fejlesztésekről.⁷

Az ilyen jellegű fejlesztőmunkába viszont be kell kapcsolódniauk történészeknek is, ugyanis a szoftverekben működő algoritmusok nagymértékben befolyásolják a keresések, adatbányászatok eredményét, így reális veszélye van annak, hogy kellő koncepcionális kontroll nélkül a technológia túlzott vagy ellenőrizetlen befolyást gyakorolhat a kutatási módszerre.⁸

DIGITÁLIS ADATARCHÍVUMOK

A digitális világ, az adatbázisok új távlatokat is nyitnak a történettudomány – de nyugodtan fogalmazhatunk általánosabban: a tudományok – számára. Egyetlen tudomány sem lehet pusztán nemzeti. Ez a tény a természettudományok esetében mindenki számára nyilvánvaló, de könnyen belátható, hogy az egyes országok története is összefügg a környező országokéval, és el kell helyeznünk saját történelmünket a világtörténelemben is.

Digitális „történelmeink” összekapcsolása nemcsak az ismeretterjesztésnek, de a kutatásnak is új távlatokat nyit. Amit „digitális történelemkönyvnek” neveztem, azt az adattárolás szempontjából nevezhetjük digitális adatarchívumnak. A digitális adatarchívumnak az a funkciója, hogy a valahol létrehozott digitális adatot,⁹ információt – amely lehet forrás, kutatási részeredmény, teljes kutatási dokumentáció forrásokkal – megőrizze, és újrafelhasználhatóvá tegye a további kutatások számára.

TÖBBNYELVŰSÉG

A nemzetközi „adatcsere” szempontjából fontos az adatbázisban tárolt dokumentumok nyelvének kérdése. Ha ugyanis azt szeretnénk, hogy Magyarország történetére vonatkozóan az internet biztosította lehetőségek alapján a hazai történettudományi műhelyek „termékei” le-

6 Az elérhető szövegelemző szoftverek részletes összehasonlító elemzését végezte el Melina Alexa–Cornelia Zuell: *A review of Software for Text Analysis*. Mannheim, [é. n.], Zentrum für Umfragen, Methoden und Analysen (ZUMA).

7 Az IASSIST 2002 júniusában Storrban tartott konferenciáján dr. Ken Reed és dr. Steven McEachern *The Prospect for „Intelligent Research”* címmel tartott előadásában számolt be az ausztráliai Deakin Universityn folyó ilyen irányú kutatásokról és eredményekről.

8 Eben A. Weitzman: *Software and Qualitative Research*. In *Handbook of Qualitative Research*. Szerk. Norman K. Denzin, Yvonna S. Lincoln. 2. kiad. Thousand Oaks–London–New Delhi, 2000, Sage Publications Inc., 818–817. o.

9 Ebben az értelemben adat egy történeti tanulmány, amely lehet például a „tankönyv” egy fejezete.

gyenek a mértékadóak hazánk határain kívül is, akkor biztosítani kell a fontosabb dokumentumok elérhetőségét a főbb világnyelveken. Ma ezt még döntően „emberi” fordítással tudjuk csak megvalósítani. Az egyre gyorsabban növekvő digitális adathalmaz több idegen nyelven történő hozzáférhetővé tételének problémája azonban az úgynevezett „gépi fordítóprogramok” fejlesztésének kérdését vetette fel. Még az angol vagy amerikai kutató számára is problémát jelenthet, ha munkájához történetesen egy francia nyelvű forrásra van szüksége, és nem tud franciául. A világhálóra felkerülő publikációk, a hozzáférhetővé váló adatbázisok az adott nyelvet nem ismerők számára lényegében nem elérhetők. A fordítóprogramok fejlesztése tehát az egész webet használó társadalom igényét szolgálja. Ezért támogatják az ilyen irányú fejlesztéseket az Európai Unió kutatásfejlesztési keretprogramjai is. A nagyobb európai nyelvek esetén már szép eredményeket értek el. Az angol, német, spanyol, francia és portugál nyelvek közötti fordításokat segítő programok¹⁰ már elérhetően hozzáférhetők, és jöllehet ellenőrzés nélkül nem alkalmazhatók, képesek akár szövegszerkesztőben megírt szövegek vagy webes böngészőbe behívott internetes oldalak tartalmának lefordítására. Rövid távon a terjedelmesebb magyar nyelvű dokumentumok gépi fordítása nem látszik megoldhatónak. Az egy-két mondatos, egységes témakörű képaláírások esetén azonban alkalmazhatónak tűnik a fejlesztés alatt álló technológia.¹¹

Az internet „egyszerű” felhasználójának természetes igénye az, hogy ha keres valamit, akkor lehetőleg egyszer kelljen beírni „valahova” a keresőkifejezést, és a „gép” hozza le az összes jó találatot. Jelenleg ez még nem működik igazán jól. Valamelyik nagy keresőrendszert használva a találatok lényegében esetlegések: vagy kezelhetetlen mennyiségű információt kapunk, vagy éppen a releváns találatok vesznek el. A probléma leglényegesebb oka talán az, hogy az egyes témakörökben még nincsenek autentikus tartalomszolgáltatók, vagy csak részlegesen érhető el az interneten, ezzel szemben nagyon sok a „szemét” az elérhető webes oldalak között. Nem látható továbbá, hogy az egy-két univerzális keresőprogram képes lesz-e nagyon általános és ezzel egy időben akár nagyon speciális tartalmak számontartására, indexelésére. Lehet, hogy az egyes területeken kialakul valamiféle „témakörököt átfogó” portál, amelyet kutatója „felkeres” a kérdéseivel. További kérdés, hogy kellenek-e nagyobb „digitális archívumok”, vagy a digitális adatokat létrehozó szervezetek alkossanak egy közös virtuális digitális archívumot, amely egy közös portálon keresztül érhető el.

A CSEREFORMÁTUMOK (SZABVÁNYOK ÉS AJÁNLÁSOK)

Bárhogyan is oldják meg ezeket a problémákat, egyre sürgetőbbben vetődik fel az a kérdés, hogy létrehozott digitális állományainkat valamiképpen egységesíteni kell, hogy más adatbázis-kezelő rendszert alkalmazó csoportok számára is használhatók legyenek, valamint azért, hogy az internet által létrejövő virtuális digitális archívumban egységesen lehessen „látni” az adatokat, azaz a pontos keresőkifejezésekkel fel lehessen kutatni a releváns találatokat. Ez az egységesítés nem jelenti az egyes intézmények önállóságának megszorítását adatbázisaik létrehozásában, azok szerkezetének, funkcionalitásának szabad megtervezését illetően. Termé-

10 Ilyen program például a Power Translator, amelyet a Lernout & Hespice fejleszt. Hasonlóan nyelvi technológiák fejlesztésével foglalkozik a müncheni székhelyű Sailabs csoport, amely például az European Visual Archive elnevezésű projekt számára fejlesztett internetes gépi fordítórendszert.

11 Magyarországon a Morphologic Kft. foglalkozik fordítóprogramok fejlesztésével, egy-két mondatos szövegek angolra, illetve angolról magyarra fordításával már szép eredményeket értek el.

szetes kívánalom, hogy egy történeti intézetben egy folyóiratcikk, bibliográfiai tétel vagy fotódokumentum tartalmi feltárása jóval részletesebb legyen, mint például egy könyvtárban vagy egy hírügynökségnél a fotók esetében. Fontos azonban, hogy legyenek olyan „elemek”, amelyek alapján valamennyi hasonló típusú dokumentum egységesen kezelhető. A kérdésnek két lényeges aspektusa van.

Az egyik a dokumentum leírására szolgáló adatelemek meghatározása. Arról van szó, hogy definiáljuk azokat a dokumentációs, tartalmi, technikai, jogi, hozzáférési adatelemeket (cím, szerző, létrehozó, a létrehozás dátuma, szerzői jogok stb.), amelyeket megadhatunk az egyes dokumentumok leírása esetén. Ezek között vannak kötelezők, ajánlottak és opcionálisan megadhatók. Több nemzetközi szervezet (DDI, Dublin Core) kötelezte magát ilyen szabványok kidolgozására. Jóllehet ezek ma még nem szabványok, csak ajánlások, a kutatói és fejlesztői társadalom idejében felismerte ennek fontosságát, és együttműködés van a digitális archívumok és a digitális tartalmakat előállító intézmények között.¹²

A másik fontos kérdés az adatelemeket és magát a dokumentumot is tartalmazó digitális adatformátum. Az egyes intézmények leggyakrabban valamilyen adatbázis-kezelőben tárolják, dolgozzák fel adataikat. A web böngészők által megjelenített HTML továbbfejlesztéseként létrehozott XML (*Extensible Markup Language*) nyelv, illetve annak folyamatos fejlesztései látszanak a legtöbb érintett szerint a legalkalmasabbnak az adatcserék és az egységes lekérdező rendszerek szempontjából a feladat megvalósítására.

Az XML nyelvben nincsen előre definiált elemlista, mint például a HTML nyelvben. A felhasználó az alkalmazásának megfelelően „beszédese” nevű elemeket használhat, amelyek „viselkedését” természetesen majd a speciális XML jelölésekhez rendelt további „felületek” közbeiktatásával kell megoldani. Az XML ezért megfelelő a relációs adatbázisok csereformátumaként.¹³

XML-alapon működik az Európai Unió 4. keretprogramja Telematikai Alkalmazások alprogramjának keretében kifejlesztett NESSTAR (*Networked Social Science Tools and Resources*) rendszer. Angol, norvég, dán, ír és spanyol digitális archiválással foglalkozó intézmények, valamint informatikai fejlesztőműhelyek hozták létre, hogy az egyre inkább digitális formában keletkező kutatási eredményeket, részeredményeket, dokumentumokat egy közös felületen az interneten keresztül elérhetővé, újrafelhasználhatóvá lehessen tenni, valamint hogy a rendszer keresőfelületén (*NESSTAR Explorer*) el lehessen érni az e rendszerben publikált valamennyi adatot és dokumentumot. A rendszer tehát köztes és végfelhasználói eszköz. Nem foglalkozik azzal, hogy az adatok milyen formában vannak tárolva, adatbázisban vagy csak fájlrendszerben. Eszközöket ad (*File Import Utility, XML Generator*), amelyek alkalmazhatóak DBASE (.dbf), Statistica (.dta, .nsf), SPSS (.por, .sav) fájlokra. Ezek előállítják a szükséges XML-fájlokat, de természetesen mi magunk is gondoskodhatunk ezek előállításáról, ami speciálisabb adatbázisok esetén elkerülhetetlen. A NESSTAR tehát az adatok interneten történő terjesztésének, publikálásának infrastruktúrája.

12 Az UK Dataarchive által 2001 szeptemberében Colchesterben *Qualitative Data: Applying the Data Documentation Initiative (DDI)* címmel rendezett workshopon az Európa valamennyi nagyobb digitális adatarchívumából jelen lévő kutatók szorosabb együttműködést kezdeményeztek újabb, a kvalitatív kutatásokhoz szükséges adatelemek meghatározására, és azoknak a DDI-be történő integrálására.

13 Az XML-lel kapcsolatban jó bevezetést ad Neil Bradley: *Az XML-kézikönyv*. Bicske, 2000, Szak.

ADATBÁZISOK ÉPÍTÉSE, ADATBÁZIS-FEJLESZTÉSEK AZ 1956-OS INTÉZETBEN

A hatvanas évek kutatásával kapcsolatban tovább folyik az intézet jelenkor-történeti adatbázisának építése. Történeti digitális archívumunk egyrészt alapanyagot szolgáltat kiadványaink (könyv, CD-ROM, internet) elkészítéséhez, másrészt segítséget nyújt további kutatásokhoz a gyorsan fejlődő informatikai eszköztár segítségével.

Fotóadatbázisunk ez ideig közel kétezer digitalizált fotódokumentummal gyarapodott. A különböző magán- és közgyűjteményekből származó fotók tartalmi és dokumentációs leírását adatbázisba töltjük. Elkészítjük a fotódokumentumok nagy felbontású, nyomdai publikálásra is alkalmas, nézőképes (600×600 pixeles keretbe illeszkedő) és bélyegkép nagyságú, digitalizált fájljait. A nagy felbontású fájl kivételével a fájlokat is adatbázisba töltjük.

Az életútinterjú-kivonatokat tartalmazó adatbázisunk teljes átszervezését tervezzük. Megkezdtük régebbi életútinterjúink rendelkezésre álló hangfelvételeinek digitalizálását (ezzel a hangfelvételek további romlását is megelőzhetjük), és a digitalizált fájlokat szintén betöltjük majd adatbázisunkba. Nemcsak a hangfelvételek, hanem a teljes szövegek digitalizálását is tervezzük. A teljes szövegek közül az 1990–1991 után készültek már megvannak valamilyen szövegszerkesztő (XY Writer, Irnok, MS Word.x, MS Winword.x) által előállított változatban. A régebbi interjúk digitalizálása azonban elkerülhetetlennek látszik. Itt alkalmazhatunk majd valamilyen OCR (karakterfelismerő) programot, de egyes rosszabb minőségű gépirások esetén mérlegelni kell az újragépelést is. 1992-ben, mikor kialakítottuk életútinterjú-kivonatokat tartalmazó adatbázisunkat, még nem volt reális lehetőség arra, hogy az interjúk teljes szövegét betöltsük egy adatbázis-kezelő programba. Sem a merevlemezek tárhelykapacitása, de főként az adatbázis-kezelők lehetőségei nem biztosították több megabájtos állományok biztonságos tárolását. Ma azonban a merevlemezek, a processzorteljesítmények és az adatbázis-kezelők fejlődésével már nincs akadálya a nagyobb, akár több gigabájtos állományok tárolásának.

Digitalizáljuk a kutatások során előkerülő fontosabb forrásokat is. Ezek főként szöveges dokumentumok, táblázatok, de lehetnek akár műalkotások, hangfelvételek, filmhíradó-részletek is. Szöveges dokumentumoknál ez főként beírást jelent, egyes esetekben szükség lehet képi megőrkítésre is, a multimédia-dokumentumokból pedig (hangfelvétel vagy videó) digitalizált fájlokat készítünk.¹⁴ A források feldolgozásakor a kutatók különböző adatokra, összefüggésekre bukkannak. A kronológiai adatokkal kiegészítik, pontosítják adatbázisunkat, megadva mindenhol az adatpontosítás forrását, azzal a kronológiai adat rekordját össze is kapcsolva.

Tovább épül bibliográfiai adatbázisunk. Feladatunknak tekintjük a korszak társadalomtörténetével foglalkozó könyvek és cikkek összegyűjtését, valamint azok bibliográfiai feltárását.

Az 1956-os forradalom utáni megtorlás során lefolytatott perekben 20–25 ezer ember volt érintett. A periratok feltárása nem csak az 1956-os forradalom és az azt követő megtorlás megértéséhez fontos. A perekben szereplő személyek adattára nélkülözhetetlen segédeszköze a jelenkori történeti és szociológiai kutatásoknak. A periratokból bőséges és nagyon részletes adatok nyerhetők a vádlottak származására, családi és vagyoni helyzetére, iskolázottságára, a forradalom előtti, alatti és azt követő életére vonatkozóan. A periratok feltárása során születő adatbázis forrásként szolgálhat a hatvanas évek társadalmának kutatásához is.

14 Minthogy a mozgóképek jó minőségű archiválása nagyon nagy, több gigabájtnagyságú fájlokat eredményez, az adatbázisba ilyenkor egy „butított”, gyengébb minőségű változatot töltünk, az eredeti, nagyon jó minőségű változatot pedig CD-ROM-on, vagy egyre inkább DVD-ROM-on tároljuk.

SZEMLE

UNGVÁRY KRISZTIÁN

Háború – élelmezés – népiptás

A holokausztkutatás legújabb eredményei Németországban

Götz Aly–Susanne Heim: *Vordenker der Vernichtung. Auschwitz und die Pläne für eine neue europäische Ordnung*; Götz Aly: „*Endlösung*”. *Völkerverschiebung und der Mord an den europäischen Juden*; Christian Gerlach: *Krieg, Ernährung, Völkermord. Forschungen zur deutschen Vernichtungspolitik im zweiten Weltkrieg*; uő: *Kalkulierte Morde. Die deutsche Wirtschafts- und Vernichtungspolitik in Weissrussland. 1941 bis 1944*; Götz Aly–Christian Gerlach: *Das letzte Kapitel. Der Mord an den ungarischen Juden*

STANDEISKY ÉVA

Költészet a Rákosi-korszakban és az 1956-os forradalom idején

Munkás, paraszt, értelmiség munkaverseny lázában ég! Agitatív antológiaköltészet Magyarországon, 1945–1956; *Piros a vér a pesti utcán. Az 1956-os forradalom versei és gúnyiratai*

SZAKOLCZAI ÁTILA

Feledendő emlék

Kisfogház Emlékhely. 2001. november 4.

EÖRSI LÁSZLÓ

Koncepciók

Jobbágyi Gábor: *A néma talp (Tóth Ilona, az orvosi kar mártírja)*

KENEDI JÁNOS

Komparatív szpionisztika

Karol Sauerland: *Harminc ezüst. Beszűgások és árulások*

KOZÁK GYULA

A Fortinbras & Fortinbras Koronaszállító Rt.

Petőcz György: *Csak a narancs volt*; Dr. Kende Péter: *A Viktor*

TISCHLER JÁNOS

A lengyel rendszerváltás – 1989

Polska 1986–1989: koniec systemu. (Referaty; dyskusja; dokumenty)

GERMUSKA PÁL

Válságok és megoldások

Sixth International Conference on Urban History. Power, Knowledge and Society in the City