

Hatszáz ezer különleges könyvtári dokumentum és a Könyvtár 2.0 – Egy újszerű szolgáltatási forma terve a nemzeti könyvtárban

Ma már nem szokatlan dolog, hogy a könyvtárak korszerű szolgáltatásokkal színesítik a kínálatukat: megjelennek a közösségi oldalakon, blogot írnak, Skype üzenetekben fogadják a referenzkérdéseket, sms-ben értesítik az olvasót az újdonságokról, hogy csak néhány példát említsünk. Ma már a Könyvtár 2.0 nem ismeretlen fogalom, hiszen a gyűjtemények weboldalain utat tör magának a „web2-es” szemlélet. Mégis, első hallásra merész döntésnek tűnik, hogy a nemzeti könyvtár úgy teszi elérhetővé gyűjteményének egy szeletét, nevezetesen több százezer digitalizált gyászjelentést, hogy a feltáró munkába bevonja a dokumentum iránt érdeklődő kutatókat, olvasókat.

Előszó

Az Országos Széchényi Könyvtár (OSZK) milliós nagyságrendben őriz és szolgáltat kisnyomtatványokat, többek között gyászjelentéseket is, melyből eddig közel 600 000 darab digitalizálása készült el egy külföldi partner közreműködésével. Mielőbb szeretnénk ezt a különleges és értékes anyagot elérhetővé tenni a felhasználók számára, de ehhez az kell, hogy a kereshetőséget lehetővé tevő leíró adatok is elkészüljenek a dokumentumokhoz. Jelenleg még cédulakatalógus formájában sem áll rendelkezésre a szükséges információ, ezért előttünk áll az a nem kis feladat, hogy a dokumentumokat feltárjuk, és az adatokat elektronikusan feldolgozzuk, adatbázisba szervezzük. Figyelembe véve a digitális állomány mennyiségét és azt a tényt, hogy a fizikai gyűjtemény évekig tartó rendezése eddig is túlzott terheket rótt az OSZK kis létszámú Plakát- és Kisnyomtatványtárának munkatársaira, meglehetősen kevés esélyt látunk arra, hogy ezt saját erőből belátható időn belül megoldjuk. Tekintettel az olvasói/kutatói igényekre és a szűkös kapacitásra, úgy döntöttünk, hogy a nemzeti könyvtár weboldalain létrehozunk egy „közösségi” szolgáltatást, és így próbáljuk meg felgyorsítani a feltáró munkát, hogy az érdeklődők mielőbb használatba vehessék a gyűjteményt.

Tudomásunk szerint még nem volt arra példa, hogy egy értékes gyűjtemény feltárásának és elérhetőségének kialakításában a könyvtárosokon kívül az érdeklődő könyvtárhasználók is részt vehessenek, pedig leginkább az ő érdekük, hogy minél több feltárt információ között kereshessenek.

A szakemberek mindig tartanak attól, hogy „laikusokat” vonjanak be a szakmai munkába. De vajon laikusnak tekintendő-e az, aki kutatói szinten, vagy akár hobbiként évek óta foglalkozik az adott dokumentumtípus kutatásával? A családfakutatók sok esetben közösségekbe tömörülnek, ahol megosztják egymással tudásukat, tapasztalatukat. Ezeket a csoportokat szeretnénk felkérni a közös munkára, a gyűjtemény feltárása és hasznosítása érdekében. Hiszünk abban, hogy akik egymás között megosztják a tudásukat, hajlandók a közös tudás gyarapítására is, vagyis saját kutatói munkájuk megkönnyítése érdekében szívesen részt vesznek az adatok javításában, rendezésében, feltárásában.

Bevezető

Köztudott, hogy a digitalizálás számos problémát vet föl: az állományvédelmi, megőrzési és technológiai szempontokon kívül a leíró adatok, a szolgáltatás és az újrahasznosíthatóság lehetőségeit is vizsgálnunk kell a projektek tervezése során.

A nagy tömegű digitalizálás technikailag ma már nem okoz igazán gondot, mivel egyre fejlődnek, specializálódnak az eszközök. Az adagolós szkennernek villámgyorsan beolvassák a lapokra szedhető könyveket, a képeslapokat és más egylapos, kisebb méretű dokumentumokat.

A gép önmagában persze akkor sem tud eredményesen dolgozni, ha egy sor műveletet automatizáltak rajta a fejlesztők. Mindig szükség van digita-

lizáló munkatársakra is, akik elvégzik a teszteket, gondoskodnak a megfelelő beállításról, a digitalizálendő dokumentum előkészítéséről, a beolvasás felügyeletéről. A digitalizáló munkaerő beállítása nem jelent komoly problémát, ha az előkészítést és a rendszeres ellenőrzést hozzáértő szakemberre bízuk. Több digitalizáló gépet is üzemeltethet egy-két betanított munkatárs, ha előzetesen alaposan megterveztük a folyamatot. (Természetesen ez nem igaz a védett, sérülékeny, vagy más okból speciálisan kezelendő dokumentumokra, amelyeket adagoló gépeken nem tudunk digitalizálni.)

Mindezekből következik, hogy megfelelő anyagi és technikai háttérrel a digitális képfájlok rövid idő alatt nagy mennyiségben készülhetnek el, de ekkor még csak egy mesterpéldányunk van, amely ebben a formában nem igazán alkalmas arra, hogy eljusson az olvasóhoz. Képkonverziós eljárásokkal csoportosan is megoldható a nagyméretű képekből a könnyebben kezelhető és megjeleníthető formátum elkészítése. Ha könyvoldalak szkennelt képeivel dolgozunk, a digitális képeket PDF fájlba szerkeszthetjük a szolgáltatáshoz. Ez a legegyszerűbb és leggyorsabb módja annak, hogy egy könyv elektronikus formában is eljusson az érdeklődőkhöz. Ha a PDF fájlokban szövegrészleteket szeretnénk keresni, vagy a szöveges oldalakat szerkeszthető, kereshető formában is elérhetővé akarjuk tenni, akkor már bonyolódik a helyzet, szükség lesz a szöveg szoftveres felismertetésére és esetleg korrektúrázására is.

Ha az eredeti dokumentum csak képi információt tartalmaz, vagy csak képi információt akarunk vagy tudunk kinyerni belőle, valamivel egyszerűbb (?) a dolgunk, mert a gondos szkennelés és utólagos képfeldolgozás (körbevágás, kontrasztállítás, esetleg retusálás stb.) után már csak a kisebb, szolgáltatható formátumot kell előállítani ahhoz, hogy a kép könnyen megjeleníthető legyen egy távoli számítógépen.

Még nem beszéltünk azonban a leíró adatokról, amelyek nélkül a mégoly gyorsan gyarapodó archívum is nehezen kezelhető digitális fájlhalmaz marad, hiszen a fájlokat nem tudjuk azonosítani, visszakereshetővé tenni. Az esetek többségében itt jönnek a gondok és a megválaszolandó kérdések:

- rendelkezésünkre áll-e egyáltalán valamilyen adat az egyes dokumentumokról?
- ha igen, az adatok milyen formában léteznek: cédulakatalógusban vagy adatbázisban?

- a leíró adatok milyen részletesek? egy konkrét dokumentumhoz kötődnek, vagy azok egy csoportjához?
- használható-e, elegendő-e a digitális fájlok azonosításához és osztályozásához a jelenleg megtalálható adatok tartalma?
- esetleg most kell feltárni az anyagot?

Mindez persze messze nem a teljes problémakör feltérképezése, csak egy áttekintő előkészítés ahhoz, amiről a jelen írás szól: egy zökkenőkel tarkított projekt tanulságairól és egy újszerű szolgáltatásra vonatkozó döntés előzményeiről. Mindenesetre egyre gyakrabban fordul elő, hogy nagyon gyorsan, nagyon sok, új digitális dokumentum áll elő, de komoly gondot okoz az azonosító adatok kérdése, a dokumentumok kereshetőségének megoldása, pedig ebben rejlik az elektronikus dokumentumok használhatóságának egyik kulcsa.

Eddig főként általános tapasztalatokról beszéltünk, most nézzük a konkrét esetet...

Az Országos Széchényi Könyvtár gyászjelentés-gyűjteménye

A széles közönség számára digitálisan szolgáltatni kívánt gyászjelentések jelenleg az OSZK egyik legszerteágazóbb gyűjtőkörével rendelkező és legtöbb dokumentumot őrző különgyűjteményi tárában, a Plakát- és Kisnyomtatványtárban találhatók meg. Tárunkban az 1711 után Magyarországon megjelent, vagy külföldön kiadott magyar vonatkozású kisnyomtatványok (azaz legfeljebb 16 oldal terjedelmű nyomtatványok) példányszáma megközelíti a 4 milliót.

Az OSZK méltán híres gyászjelentés-gyűjteményének projektszintű könyvtári feldolgozása és kutathatóvá tétele két évvel ezelőtt, 2007 augusztusában zárult le. Hosszú és fáradságos út vezetett odáig, hogy az online szolgáltatás gondolata reálisan felmerülhetett.

A gyászjelentés-gyűjtemény átfogó és tervszerű rendezésének hátterében az ún. „mormon program” állt. A genealógiai adatokat világszerte módszeresen gyűjtő, megőrző és szolgáltató *GSU (Genealogical Society of Utah)*, amelyet a Salt Lake City-ben (Utah állam, USA) székelő *Mormon Egyház* működtet, 2003 januárjában kötött megállapodást az OSZK-val a gyászjelentés-gyűjtemény mikrofilmzéséről és digitalizálásáról. A megállapodás a következő fontos részleteket tartalmazta:

egyik részről a dokumentumokat őrző Plakát- és Kisnyomtatványtár (akkori nevén Kisnyomtatványtár) vállalta, hogy a 2002. december 31-éig megjelent gyászjelentéseket mikrofilmmezés céljából rendelkezésre bocsátja; a GSU pedig kötelezte magát arra, hogy fedezi az OSZK által elkészített felvételek költségét, továbbá vállalta a negatívok kidolgozását és azt is, hogy a pozitív mikrofilmekből és azok digitális változatából egy példányt térítésmentesen átad az OSZK-nak.

A projekt megvalósulásának alapvető feltétele a gyászjelentés-gyűjtemény rendezése, revíziója és feldolgozása volt. E munkának minden nehézsége és súlya a Plakát- és Kisnyomtatványtár jelenlegi és a már nyugdíjba vonult munkatársait terhelte. A korábban több részgyűjteményből álló anyagot, amely mint utóbb kiderült, 533 000 gyászjelentést foglalt magába, egységes elvek szerint és új szabályzat alapján, szigorú betűrendbe rendezték a kollégák. Ez a munka a nagyalakú kéziratos és a XVIII–XIX. századi szomorúlevelek, szomorújelentések leválasztása után kezdődhetett el. A rendezés során több szempontot kellett figyelembe venni, például a ragadványneveket, vagy az azonos nevűek megkülönböztetését. A nyomdai kiadásváltozatokat és a nyelvi változatokat is el kellett különíteni. Az egyik nagyon fontos munkafolyamat a duplumok azonosítása volt, hiszen a fotózáshoz számos esetben ki kellett választani a kutathatóság szempontjából legjobb minőségű példányt.

A betűrendezett gyászjelentések végleges, beszámozott palliumokba kerültek, amelyekre a dokumentumokon szereplő családneveket is rávezették. Ezután listák készültek a restaurálásra való előkészítés, illetve fotózásra átadás céljából. A statisztikai kimutatások szerint a majdnem öt éves munka során az OSZK Restauráló Osztályán 2007 tavaszáig 11 543 (!) gyászjelentést 20 147 (!) munkaóránban restauráltak. Az OSZK keretein belül működő korábbi két restauráló műhely munkatársai a rendezés és előkészítés során feszített tempóban végezték el ezt a hatalmas munkát.

A tár dolgozói a többszázszoros gyűjteményt nagyjából tízezres adagokban és számozott dobozokban adták át a Fotólabor számára, a neveket tartalmazó listával együtt. A mikrofilmek elkészítése után, az adott tekercs számát tüntették fel a listán szereplő családnevek mellett. E listákat a későbbiekben a mikrofilmeket őrző és szolgáltató Mikrofilmtár is megkapta, így lehetővé vált a családnevek kutatása a mikrofilmeken.

A tár legkutatottabb dokumentumai már a mikrofilmek elkészítése előtt is a gyászjelentések voltak. A hagyományos könyvtári dokumentumnak nem nevezhető, egylapos, ritkábban többlapos nyomtatványok számos értékes adatot és információt tartalmaznak. A szomorúlevelek, szomorújelentések és a gyászjelentések a legkülönbözőbb szakterületek kutatói számára szolgálnak értékes forrással. Első helyen kell megemlíteni a családfakutatókat és a biográfusokat, de a korszakra specializálódott történészek és szociológusok is jól használhatják e gyűjtemény darabjait. A nyelvtudománnyal és a nyelvtörténettel foglalkozók számára elsősorban a korábbi, XVIII. századi szomorújelentések érdekesek. A művészettörténészek e dokumentumokat formai szempontból kutathatják, hiszen a gyászjelentések gyakran tartalmaznak grafikus elemeket, vagy grafikai alkotásokat, továbbá iniciálét, díszítő motívumokat. A különböző felekezetek elhunytjai számára készített gyászjelentések az eltérő vallási jelképek használatára vonatkozóan szolgáltatnak adatokat. A nyomdatörténészek, sőt a vízjelkutatók is gazdag forrásanyagot találnak saját kutatási területükhöz a gyászjelentés-gyűjtemény egyes részeiben.

Az itt felsorolt kutatási szempontok közül természetesen a „mormon program” végső célja nem más volt, mint genealógiai adatok összegyűjtése és hozzáférhetővé tétele. Ebből a szempontból kiemelendők az adatokban gazdagabb és formalizáltabb XIX–XX. századi gyászjelentések. Ezek elődei, a XVIII. századi szomorújelentések, formailag kevésbé voltak kötöttek, és nem egyszer irodalmi alkotásként is megállták a helyüket, de a családfakutatók számára jóval kevesebb adatot tartalmaznak.

Bár eltérő struktúra alapján, de a későbbi gyászjelentéseken feltüntetették az elhunyt neve mellett az illető halálának és temetésének dátumát, helyét, az elhunyt életkorát, betöltött közeleti funkcióit, hivatalos elismeréseit. Megtaláljuk a családtagok felsorolását, akiknek rokoni kapcsolatai nem csupán az elhunyttal, hanem egymással is jól megállapíthatók. Az adatok eddig még nem ismert rokoni kapcsolatokra deríthetnek fényt. Életrajzírók, lexikoncikkírók ellenőrizhetik, és esetleg javíthatják a híres személy hosszú időn keresztül tévesen használt adatait. Hosszan folytatható lenne még a sor, hogy e jelentős gyűjtemény milyen területeken nyújthat biztos támpontot és szolgálhat új információval. Bízunk benne, hogy az OSZK-ban – az Elektronikus Dokumentum Központ és az Informatikai Igazgatóság munkatársai által – az utóbbi

hónapokban fejlesztett szolgáltatás könnyebb hozzáférést nyújt majd a kutatók és az olvasók számára, valamint lehetővé teszi, hogy a Plakát- és Kisnyomtatványtárban található gyászjelentés-gyűjtemény az interneten keresztül közkinccsé váljon.

Anélkül, hogy a megvalósuló online szolgáltatás és a mögötte található, mára jól rendezett fizikai gyűjtemény jelentőségét kisebbíteni szeretnénk, bizony le kell vonnunk a tárban elvégzett, közel öt éves előkészítő munka negatív tanulságait is.

Ahogy azt e tanulmány az ún. „mormon-projekt”-tel kapcsolatban leírja, az nem önzetlen segítségnyújtás volt az OSZK számára, habár a végeredmény kétségtelenül hasznos és látványos. A mára birtokunkban lévő hatalmas digitális gyűjteményért súlyos árat fizetett a nemzeti könyvtár, a Plakát- és Kisnyomtatványtár és annak egyes munkatársai is. Ahhoz, hogy az egész fizikai gyűjtemény teljes mikrofilmzése elkezdődhessen és megvalósulhasson, a tár akkori vezetése a Plakát- és Kisnyomtatványtár alapfeladatainak elvégzését a minimumra csökkentette. Ennek az lett a következménye, hogy a gyarapodás során a tárba érkező, közel évi 30 000 dokumentum szétválogatása, alapszintű feldolgozása lényegében több mint négy évre leállt. A gyászjelentések szigorú rendezésével, feldolgozásával és fotózásra előkészítésével töltött évek alatt a felgyülemlett dokumentumok száma megközelítette a 120 000-et. Ez a „be nem kalkulált” lemaradás a tárat mostanáig sújtja, és napi szinten többletmunkát ró a munkatársakra. Emellett a többi részgyűjtemény korábban folyamatosan végzett feldolgozása is leállt (gondoltunk akár a régi metszetekre, grafikai plakátokra, XVIII–XIX. századi nyomtatványainkra stb.). Hangsúlyozni kívánjuk, hogy a részgyűjtemények projektszerű feldolgozása ellen – ahogy ez a gyászjelentésekkel történt – a tár mai vezetésének nincs kifogása, hiszen a számtalan részgyűjtemény párhuzamos feldolgozása talán kevésbé hatékony, mint az egy-egyféle dokumentumtípusra koncentrált munka. Ugyanakkor a „mormon-projekt”-tel induló és napjainkban is folyó gyűjteménydigitalizálás tervezésekor nem hagyhatunk ki egy rendkívül fontos tényezőt: a digitális gyűjtemény kizárólag egy már könyvtári szempontok alapján rendezett gyűjteményről történő digitalizálással készülhet el. E rendezés és feldolgozó munka emberi erőforrást igényel, amellyel szakmailag megalapozottan és átgondoltan szükséges gazdálkodni.

A „mormon-projekt”

Amint azt a gyűjtemény ismertetőjében is olvashattuk, a nemzeti könyvtár a Mormon Egyház (innen a projekt neve) kereste meg azzal a javaslattal, hogy ingyen elkészítené az OSZK-ban található gyászjelentések digitális másolatát, ha mikrofilmen megkaphatja a dokumentumokat. Persze, nem önzetlen segítségnyújtásról volt szó, hiszen saját honlapjukon (<http://pilot.familysearch.org>) kívánták szolgáltatni a képeket, elsősorban a családfakutatók érdeklődésére számítva. A projekt során elkészült digitális képeket természetesen a nemzeti könyvtár is szolgáltathatja, vagy bármi módon hasznosíthatja, korlátozás nélkül. Az OSZK egyébként minden esetben ragaszkodik ahhoz, hogy a könyvtárban (akár külső megrendelésre) történő digitalizálás során létrejött elektronikus dokumentumokból egy példány az ő gyűjteményébe kerüljön.

Az OSZK-ban elkészültek a mikrofilmek, amelyeket átadtunk digitalizálásra a partnercégnek. Ezt követően kezdődtek a gondok: a munka meglehetősen lassan haladt, a kapcsolattartás akadozott, ígéreteket ugyan kaptunk néha, de a digitalizált anyagokat sokáig nem láttuk. (Az első minták 2007-ben érkeztek.) Leveleinkre nem, vagy csak jelentős késéssel jött válasz. Időnként már csaknem lemondtunk arról, hogy valaha elkészülnek a szkennelések. Többszöri egyezkedés után (és több éves csúszással) végre megérkezett hat darab hordozható tárolóeszköz a digitális fájlokkal. Az eszközöknek speciális csatlakozójuk volt, de találtunk megfelelő összekötő kábelt, hogy belemézhesünk a lemezek tartalmába.

A kezdeti problémák

Megnézve az állományokat kiderült, hogy nem pont olyan minőségben készültek, mint amilyet vártunk, például nem voltak körbevágva a képek, vagy ferdén helyezkedtek el a háttérként alátett lapon (ez utóbbi a felvételek készítésénél igényelt volna nagyobb figyelmet). Igaz, az átadás feltételeit a két fél között létrejött szerződés nem részletezte kellő mértékben. A megállapodás kitért a felhasználás lehetőségeire, de az átadás minőségi feltételeit nem írta körül megfelelő részletességgel, így reklamálni is nehéz lett volna.

A megállapodás idején a mikrofilmzés úgy történt, hogy az egyes gyászjelentéseket egy A4-es, fekete lapra helyezték, és előfordult, hogy kis cédulákat is tettek a háttérlapra további információkkal, pél-

dául „color printed” felirattal. Döntés kérdése, hogy eldobjuk vagy megtartjuk ezt a háttér és a rajta lévő információkat. Vitathatatlan, hogy „elegánsabb” lenne, ha maga a dokumentum jelenne meg a képernyőn a fölösleges háttér nélkül. Létezik megoldás a hasznos terület automatikus körbevágására, a fölösleges háttér lebegésére, de a fenti körülményekre tekintettel ezt nem is olyan egyszerű megoldani. Gondoljunk csak arra, hogyan is néz ki egy gyászjelentés? Általában a szélein fekete keret van... ha a háttér is fekete, bizony elég nehéz automatizáltan (sőt manuálisan is) beazonosítani a körbevágandó méretet. Várhatóan elveszne a fekete keret is, ami ennél a dokumentumtípusnál nem igazán korrekt megoldás. Sajnos a mikrofilmezés során nem minden esetben cserélték fehérre a háttérlepet, amikor az eredeti dokumentum fekete szegélye a papír széléig kiért. Sokszor előfordul, hogy ezen a szegélyen egy minta van, például egy vékony vonalakkal rajzolt kereszt, ami adott esetben az elhunyt vallási hovatartozásáról is információt hordozhat. Ezzel vajon mit kezd az automatizált körbevágás?

Kérdés, mi az egyszerűbb: néhány próbálkozás után megközelítő pontossággal beállítani az automatikus körbevágás paramétereit, majd utána megkeresni és korrigálni a rosszul sikerült példányokat, vagy inkább egyenként beállítva, „kézzel” levágni a fölösleget, ami ilyen mennyiségénél rendkívül idő- és költségigényes megoldás. Legrosszabb esetben megtehetnénk, hogy a teljes beszkenelt képet (háttérleppal együtt) használjuk fel, és azt kicsinyítjük szolgáltatható méretre. Ez egyszerűbben automatizálható lenne, de kevésbé jól használható eredményt hoz.

Mint már jeleztük, közel 600 000 kép kezeléséről van szó! Ilyen mennyiségben a „kézi munka” rendkívül erőforrás-igényes, mi pedig nem rendelkezünk a szükséges szabad kapacitással. Anyagi forrásaink nincsenek külső segítő foglalkoztatására, így marad a gépidő optimális kihasználása és a rendelkezésünkre álló emberi erőforrás bevetése, ami egyszerűbben fogalmazva saját munkatársaink fizikai és szellemi erejét, valamint véges munkaidejét jelenti. Végül úgy döntöttünk, hogy megállapítottunk egy olyan körbevágási méretet, amelyikbe mind a fekvő, mind az álló gyászjelentések belefértek, és erre a méretre állítottuk be az automatikus körbevágást.

Térjünk vissza a digitális képekre, amelyek eredeti állapotukban 1 MB körüli méretűek, szürkeárnyaltosak és 200 dpi felbontásúak. Az OSZK-ban július

elején született meg az a belső szabályzat, amely a képi dokumentumok szolgáltatásának módjáról rendelkezik. Ennek megfelelően ezt a dokumentumcsoportot 72 dpi felbontásban szolgáltatjuk. A képeket ennek megfelelően konvertáltuk, méretüket 50%-osra csökkentettük, ezzel a helyfoglalásuk is mintegy harmadára csökkent. A szolgáltatási ajánlásnak megfelelően „vízzel” is ellátjuk a dokumentumokat. A képmanipulációkhoz léteznek olyan ingyenes szoftvereszközök, amelyekkel ez a művelet egyszerre nagyobb fájlcsoponton is elvégezhető, legfeljebb szabad gép és egy kis munkaszervezés kell hozzá. A feladathoz ingyenes képkezelő programot használtunk.

Mivel a gyászjelentések teljes tartalma fontos részinformációkkal szolgálhat, ezért szükséges a szövegfelismertetés is. A gyászjelentések esetében a teljes szövegű kereshetőségnek azért van különös jelentősége, mert kezdetben csak minimális leíró adatot tudunk szolgáltatni, a szöveg pedig a kutatók számára további hasznos információval szolgálhat. A szövegfelismertetés a nagyméretű képeken végezhető el a legjobb hatásokkal. Megfelelő folyamatkezeléssel és adminisztrációval bevonhatunk több gépet is a feldolgozásba, hiszen a gépidő jobb kihasználása érdekében éjjel is futhatnak olyan előre definiált folyamatok, amelyek nem igénylik a munkatársak jelenlétét, beavatkozását.

Az eredeti dokumentum és a digitális másolat kapcsolata

Az eddig vázoltakhoz képest nagyságrendekkel komolyabb problémát jelent a képek azonosítása és kereshetővé tétele, valamint a kapcsolódó információkat tartalmazó adatbázis kialakítása és feltöltése. Azt eldöntöttük, hogy szeretnénk az olvasók számára elérhetővé tenni a gyászjelentéseket, hiszen az érdeklődés itthon is jelentős, mind a családjuk története iránt érdeklődők, mind a különféle szakterületek kutatói körében. A jól használható szolgáltatás kialakításához még kevés, ha körbevágott, kicsinyített képeink vannak. Hogyan tesszük kereshetővé a dokumentumokat? Hogyan szolgáltatjuk a kutatók számára? Mi az a minimális adatsor, amit a képekhez kell kapcsolnunk? A szolgáltatásnak csak akkor van értelme, ha a képek egyenként azonosíthatók. Az anyag kutathatóságához további adatokra és a teljes szövegtartalom kereshetőségére is szükség van.

Amint láttuk, a mikrofilmezés előkészítése során a gyászjelentéseket az elhunyt családneve alapján ábécésorrendbe rendezték, és ebben a sorrend-

ben készültek a felvételek is. A digitális másolatok folyamatos számozással, csak sorrendi információt tartalmazó fájlnevvvel érkeztek, filmtekercsenként egy-egy mappába rendezve. Beazonosítottuk, hogy melyik mappa melyik mikrofilmet tartalmazza, és ekkor derült ki, hogy hét film anyaga hiányzik a beérkezett tárolóeszközökről. Lehet, hogy a digitalizálás során maradtak ki, de az is lehet, hogy elkészült ugyan a szkennelés, de valami miatt a hordozóra nem került föl a digitális kép. Ennek utánajárunk, remélhetőleg viszonylag egyszerűen pótolható a hiány.

Amíg a digitalizálás készült, az OSZK Plakát- és Kisnyomtatványtárának munkatársai adatbázisban rögzítették a mikrofilm azonosítóját, valamint a kezdő és záró családneveket, és elkészült a gyászjelentéseken előforduló mintegy hetvenhat-ezer (!) családnév jegyzéke, amit szintén az adatbázisban rögzítettünk, azzal a céllal, hogy a felhasználók könnyebben eligazodjanak a nyomtatott gyászjelentések között. Ezzel lehetővé vált, hogy egy családnévre keresve kiderüljön, hogy egyáltalán van-e a gyűjteményben a keresett családra vonatkozó gyászjelentés, és ha van, melyik mikrofilmen található a másolata. A digitális képek beérkezésekor ezekhez kötöttük hozzá a mappaneveket, bár ezzel is csak a mikrofilmhez kapcsolódó információhoz tudjuk irányítani az olvasót, mert az az információ még nem állt rendelkezésünkre, hogy melyik kép konkrétan melyik nyomtatott gyászjelentéshez tartozik. Ez még mindig igen szerény visszakeresési lehetőséget nyújt. A közelebbi találatokhoz az olvasónak a mappában mintegy 900-1000 képet kéne átlapozni, ami rendkívül nehézkes. Ennél mindenképpen jobb feltételeket kell teremteni, ha az interneten szabadon elérhetővé és kereshetővé, kutathatóvá kívánjuk tenni a gyűjteményt.

Az egyes képek azonosítása

Két megoldást vizsgáltunk meg. Az egyik, hogy a tár munkatársainak kell beazonosítani a képeket, vagyis szakszerűen (de legalább az azonosíthatósághoz szükséges minimális adatsorral) le kell írni a dokumentumokat. Figyelembe véve, hogy a jelenleg kb. 3,5 millió egységet tartalmazó kisnyomtatvány-gyűjtemény feltárása leköti a teljes kapacitást, továbbá, hogy évente kb. 30 000 egység érkezik, a jelenlegi munkatársak már nem terhelhetők tovább. Mivel a feladat elvégzésére nincsen semmilyen anyagi forrás, külső munkaerő bevonására sincs mód, ezért elég reménytelennek látszik a helyzet. Ráadásul szeretnénk mielőbb (terveink

szerint még 2009-ben) szolgáltatathatóvá tenni a képeket, tehát kevés idő áll rendelkezésre a feladat elvégzésére és a szolgáltatás béta verziójának elindítására.

Szövegfelismertetés

A hiányzó, részletes leíró adatok átmeneti pótlására megoldásként kínálkozik a nyers szövegfelismertetés, és a teljes szövegű keresés kialakítása. Tesztet végeztünk, hogy lássuk, milyen hatásokkal ismeri fel az OCR program a karaktereket, mennyire válnak kereshetővé a fő szempontot jelentő nevek, helynevek, évszámok. Az eredmény elég vegyes volt. A legnagyobb nehézség ebből a szempontból, hogy a kulcsfontosságú információ, vagyis az elhunyt neve, éppen a kiemelés miatt speciális tipográfiájú betűkkel szerepel a nyomtatványon, ami kétségesse teszi a név pontos azonosítását. Sok esetben pedig nem is teljes név szerepel az eredeti dokumentumon, hanem például ennyi: „testvérünk, János”. Mivel az elhunyt azonosítása kulcskérdés, elkerülhetetlen, hogy legalább az ő teljes nevét pontosan azonosítsuk, és a hozzárendeljük a konkrét dokumentumhoz. Úgy ítéltük meg, hogy a használható szolgáltatás érdekében ez mindennél fontosabb, tehát megpróbálunk erre valamilyen – viszonylagosan gyors és megbízható – megoldást találni. A tapasztalatok szerint a név azonosítását nem bízhatjuk kizárólag a gépre, egészen biztos, hogy emberi beavatkozásra is szükség lesz.

Az egyes gyászjelentések automatikusan felismertett szövege a kép fájlnevével azonos néven txt kiterjesztéssel, a képpel azonos mappába került. Így könnyebben megvalósítható, hogy a keresett gyászjelentés képe és szövege összehozható legyen, illetve a teljes szöveg kereshető legyen. Ez már kicsit jobb helyzet, de még nem tartunk ott, hogy az elhunyt nevére keresve el lehessen jutni a konkrét képfájlhoz.

Körvonalazódnak a legfontosabb feladatok

Célkitűzés:

- lehetőleg rövid idő alatt elérhetővé kell tenni a gyászjelentések digitális változatát,
- keresési lehetőségeket kell nyújtani
 - o legalább az elhunyt neve alapján,
 - o lehetőleg a teljes szövegben,
 - o optimális esetben részletesebben feltárt adat-sorban,
- kutathatóvá kell tenni a dokumentumokat.

Összegezve az eddigieket, a szolgáltatás előkészítése érdekében a következő teendők körvonala-zódnak:

1. teszt jelleggel konverziót és szövegfelismerte-tést kell végezni egy kisebb mintán, hogy fel tudjuk mérni az elvégzendő munkához szüksé-ges erőforrásigényt;
2. fel kell mérni, hogy hogyan tudunk haladni a jelenlegi kapacitásunkkal a képek átalakításával és OCR-ezésével, szükség van-e további erő-források bevonására?
3. a képekből szolgáltatási méreteket kell konver-tálni;
4. a teljes szövegű kereséshez OCR-ezni kell a dokumentumokat;
5. meg kell tervezni a szolgáltatáshoz a leíró ada-tok minimális és ideális körét;
6. a képeket egyenként azonosíthatóvá kell tenni, minimálisan az elhunyt neve alapján;
7. ki kell alakítani a keresési lehetőségeket, vala-mint egy olyan felületet, amelynek segítségével együtt jelenhet meg a gyászjelentés képe és a felismertett szövege.

1. Az első tesztek azt mutatták, hogy megoldható a tömeges szövegfelismertetés, méghozzá úgy, hogy az elkészült TXT fájl a kép mellé, azzal azonos mappába kerül, és a képfájllal megegyező nevet kap. A folyamat meglehetősen jól automatizálható, nem igényel állandó felügyeletet, tehát a felismertetés napközben is futtatható a szabad kapacitással rendelkező gépeken, illetve éjszakára elindítható néhány gépen.

2. A képkonverzió és az OCR elvégezhető saját erőforrásokkal, ha több gépen megszervezzük az éjszakai futást is. Ez utóbbi művelet az archív képfájlokat használja, melyek a belső, korlátozott hozzáférésű tárhelyen vannak, biztonsági okokból csak kijelölt gépekről és meghatározott bejelentke-zési azonosítókkal érhetők el. Ezt a korlátozást nem is szeretnénk feloldani még átmenetileg sem, így ezt a munkafolyamatot kizárólag belső munka-társak végzik: az Elektronikus Dokumentum Köz-pontban és a Plakát- és Kisnyomtatványtárban dolgozó kollégák vesznek részt a munkában.

3. A képkonverzió párhuzamosan történhet az OCR-ezéssel. A kiválasztott, ingyenes képkézelő program jó adottságai megfelelnek az igényeink-nek, egyszerre nagyon sok fájlt tud kezelni, és viszonylag gyorsan végzi el az átalakításokat. Egy menetben megtörténhet a körbevágás és a kicsi-nyítés. A szolgáltatási képeket az archív verzióktól elkülönítetten tároljuk.

4. Az OCR-ezést automatizáltan végezzük. Te-kintve a gyászjelentések sokféle betűtípusát, és a nyomtatás eltérő minőségét, a szövegfelismertetés vegyes eredménnyel jár. Arra nincs lehetőségünk, hogy a szövegeket ellenőrizzük és javítsuk, egy-előre meg kell elégednünk egy nyers szövegválto-zattal.

5. Az OSZK Plakát- és Kisnyomtatványtárának munkatársai jól ismerik a gyűjteményt és a kutatók igényeit, így meg tudják határozni a gyászjelenté-seket leíró adatok körét, és ki tudják jelölni az adatsorból azokat a mezőket, amelyek minimálisan szükségesek az egyes képek azonosításához, a gyűjtemény kereshetőségéhez. A digitalizált gyászjelentések nagy száma miatt a feltárás meg-lehetősen időigényes munkája a napi feladatok mellett jelenleg nem fér bele a kollégák munkaide-jébe.

6. A megfelelő minőségű szolgáltatás feltétele az elhunyt nevének és a gyászjelentés képének ösz-szekapcsolása, valamint felismertett nyers szö-veg megjelenítése. A kutatás támogatásához javí-tott szövegre, és adatbázisba rendezett részlete-sebb adatsorra is szükség lenne, de ennek az elkészítése még hosszú időt vesz igénybe.

7. A szolgáltatás kialakítása során még számos kérdésről kell döntenünk:

- megvárjuk-e, hogy teljesen feltárt legyen a gyűj-temény, vagy egy minimális adattal már elérhe-tővé tegyük a képeket?
- első megközelítésben megelégszünk-e az el-hunyt nevének pontos beazonosításával és az erre épülő kereséssel?
- ha igen, akkor hogyan tudjuk gyorsan és vi-szonylag hibamentesen hozzárendelni a képek-hez a teljes neveket?
- mi legyen a minimális és az optimális leíró ada-tor, amit a képekhez kell rendelni a jobb kutatha-tóság érdekében?
- hogyan jelenítsük meg a képeket, a szöveget és a rendelkezésre álló, kereshető adatokat?
- hogyan tudnánk elősegíteni, hogy mielőbb javí-tott szövegek és pontos adatok álljanak a kutatók rendelkezésére?

Hosszas egyeztetések után úgy döntöttünk, hogy a lehető leghamarabb közkinccsá tesszük a képeket, hiszen a korábbi partner is elérhetővé tette a digi-talizált állományt és a családfakutatók még ebben az egyszerű formában is nagy érdeklődést mutat-nak a gyűjtemény iránt. Bár az OSZK több infor-mációval rendelkezik a dokumentumokról, de ezek

jelenleg csak a könyvtárba betérő kutatók számára hozzáférhetők. Úgy látjuk, hogy a szerényebb szolgáltatásra is van igény, ezért több lépcsőben alakítjuk ki az interneten is elérhető keresőfelületet.

Az első lépcső, hogy az elhunyt teljes neve alapján el lehessen jutni a konkrét képhez. Kollégáink őszre kialakítanak egy olyan felületet, amelyen látható a gyászjelentés kicsinyített képe, valamint az előforduló családnevek és a lehetséges keresztnévek listája. Kattintással lehet a listákból kiválasztani a családnevet és a keresztnévet (valamint a többi névelemet), amelyek automatikusan bekerülnek abba az adatbázisba, amelyben a kicsinyített kép neve, elérési útja és a képet tartalmazó mikrofilm azonosítója is szerepel. Mivel nem lehetünk biztosak abban, hogy a névlisták tökéletesek, azért a személyneveket tartalmazó mezők a kiválasztás után még javíthatók, vagyis el lehet végezni a szükséges módosításokat, hogy lehetőleg pontos nevet tároljunk az adatbázisban. Ezzel a módszerrel viszonylag gyorsan és hibamentesen megoldható a művelet, de 600 000 kép esetében ez sem kevés munka: ha képenként fél percet számolunk, akkor is 5 000 óra munkaidőt kell rászánni. Ez öt embernek közel félévi munkája! Egy összetett név (előzékenév, nemesi előnév, rang, férjezett név, leánykori név, több keresztnév stb.) leírására adott esetben a fél perc nem is elég. A munkát tovább lassítja, ha nem azonosítható egyértelműen az elhunyt neve, mert csak ennyi szerepel a név helyén: „Anyika”, és a szövegből sem derül ki a pontos név. Ilyen esetben már komoly kutatómunkát igényel, ha pontosabb adatot szeretnénk feltüntetni az adatbázisban. Be kell látnunk, hogy erre most nem tudunk vállalkozni.

A szűkös kapacitás kiegészítésére csak egy megoldás látszik: önkéntesek bevonása, pontosabban együttműködés a gyászjelentések kutatása iránt érdeklődő közösségekkel!

Az önkéntesek nagyon sokat segíthetnek abban, hogy egyre több javított szöveg álljon rendelkezésre, és ezzel a teljes szövegű keresés találati biztonsága is növekedjen. Előkészítünk adatbázismezőket is, amelyekbe a szövegek javítása során beírható vagy átemelhető a megfelelő adat, és ezzel fokozatosan feltölthető a leíró adatbázis is. A munkát gyakorlott könyvtárosok fogják irányítani és ellenőrizni.

Már tervezzük azt a felületet, amelyen a további feltáráshoz alapvetően szükséges minimális adat-

tal, vagyis az elhunyt nevével azonosítva megjelenítjük a dokumentumot, és mellette szerkeszthetően elhelyezzük a felismertetett szöveget. A keresés eleinte csak az elhunyt nevére korlátozódik, illetve a nyers szövegben lehet keresni egy szövegrészlet alapján. Böngészési lehetőség is lesz, még hozzá az elhunyt családnevének ábécésorrendjében megjelenített listák segítségével. A nagy találati halmazokban többféle navigálási módot alakítunk ki: adott találati oldalra lépést, lapozást stb.

A tervezett adatbázis az elhunyt néven kívül tartalmaz magára az eredeti, illetve a digitális dokumentumra vonatkozó adatokat is (pl. azonosító, fájlnev, elérési útvonal, az eredeti dokumentum keletkezésére, méretére vonatkozó információk). Lehetőség szerint rögzítjük az elhalálozás és a temetés helyét és idejét, valamint az elhunyt életkorát, ha ez kideríthető. Sok esetben olvasható a gyászjelentéseken a gyászolók neve és az elhunythoz való viszonyuk, ami nagyon fontos információ lehet a kutatóknak. Ezeket a neveket is felvesszük az adatbázisba, és minősítővel látjuk el (rokonsági fok vagy egyéb kötődés), hogy ennek alapján felderíthetők legyenek a szerteágazó rokonkapcsolatok. Minél tökéletesebb a szövegjavítás, és minél pontosabbak a dokumentumról felvett adatok, annál jobban kiaknázható lesz az adatbázisban fokozatosan felépülő gazdag információhalmaz.

Megjegyzésmezőket is fenntartunk arra az esetre, ha az előre felkínált mezőkbe nem illik bele valamely fontos információ, de mégis szeretnénk azt megőrizni. Példaként említhetjük, ha egy szervezet vagy intézmény saját halottjának tekintette az elhunytat, vagy valamilyen fontos információ olvasható a tevékenységével, életével, halálával kapcsolatban.

Kérdések az újszerű szolgáltatási forma kialakítása körül

Már az előkészítő munka során, a pontos nevek hozzárendelésében is jól jönne a segítség, de mint mondtuk, ez biztonsági okokból nehezen megoldható. A felismertetett szöveg javítására és a legfontosabb adatok adatbázisba emelésére végképp nem lenne esélyünk, ha nem vennénk igénybe külső közreműködőket. Mindenképpen szeretnénk ellenőrzés alatt tartani a munkafolyamatokat, de azt is tudjuk, hogy ezek a közösségi szolgáltatások akkor működnek jól, ha egyfajta bizalomra épül-

nek. Meg kell találni az egyensúlyt a bizalom és az ellenőrzött működés között. Bízunk abban, hogy minden résztvevőnek érdeke, hogy ne sérüljenek, hanem folyamatosan javuljanak a szövegek és a leíró adatok. Vannak kapcsolataink azokkal a kutatócsoportokkal, amelyeket szívesen bevonnánk a munkába. Csak regisztrált felhasználók kapcsolódhatnak be az adatjavításba, és lehetőség szerint őket is úgy választjuk meg, hogy valamelyik fent említett kutatói közösséghez tartozzanak,

ezzel is bizonyítva érdekeltségüket az adatok minőségének megőrzésében.

Természetesen mi is izgatottan várjuk, hogy ez az elképzelés hogyan válik be a gyakorlatban, hiszen még nincs tapasztalatunk ilyen típusú szolgáltatás létrehozásában és működtetésében.

A közös munka várható veszélyei (1. táblázat)

1. táblázat

Veszélyek és elhárításuk

Veszély	Megoldási javaslat
Még nem volt hasonló szolgáltatás az OSZK-ban, nincs ilyen esetre kialakított szolgáltatás politika.	Ki kell alakítani a szolgáltatás használatának biztonságos feltételeit. Béta verzióban kell tesztelni a készülő felületet. Be kell vonni a tervezésbe a leendő közreműködőket.
Előfordulhat szándékos szövegrongálás.	Mivel együtt jelenik meg a képernyőn a szkennelt kép és a felismertetett/javított szöveg, azért könnyen észlelhető, ha a kettő között jelentős eltérés van. Véglegesen megőrizzzük a nyers OCR-ezett változatot, amihez kritikus esetben vissza lehet térni. A regisztrációval biztosítható a közreműködők azonosítása. Kitiltjuk azt a felhasználót, aki nem megfelelően kezeli az adatokat. Bízunk benne, hogy nem a rongálás a közösség célja, inkább segíteni fognak abban, hogy az ilyen eseteket felderítsük, és megfelelően kezeljük.
A felhasználók/kutatók nem szállnak be a várt mértékben a szövegjavításba.	Először moderátlanul engedjük a javításokat, hogy lássuk, mennyire aktívak a felhasználók. Megfelelő kommunikációval tudatosítjuk a közreműködőkben, hogy miért hasznos a részvételük. Bevonjuk a könyvtárakat a szolgáltatás népszerűsítésébe, mert a nagyobb használati igény jelentkezése is inspirálhatja a közreműködőket. Legrosszabb esetben a szolgáltatás megmarad az első szinten, vagyis a név alapján történő keresés eredményeként megjelenik a dokumentum képe és nyers szövege. Ezt a szintet – ha lassan is – de saját erőből mindenképpen létrehozuk.
Nem tudjuk előre felmérni, hogy milyen aktivitást tanúsítanak majd a közreműködők, és ehhez milyen szerverkapacitásra lesz szükség.	A béta verziót kisebb körben teszteljük, a technikai feltételeket egy fokozatosan növekvő használatnak megfelelően szükség szerint módosítjuk.
Nem lesz egységes a javítás, a feltárt adatsor stb.	A közreműködő csoportokat felkészítjük, hogy egységes szemlélettel végezzék a javításokat. A munka menetét szervezett keretek között igyekszünk tartani, a munkamegosztásról rendszeresen egyeztetünk. Minden elvégzett javítás egy fokkal használhatóbbá teszi az adatokat, akkor is, ha ez nem egyenletesen történik az adatbázis egésze szempontjából.
Ha beindítjuk a közösségi szövegjavítást, ez feladatokat jelent a könyvtárosoknak is, de még nem tudjuk megítélni, mennyi kapacitást köt le (pl. technikai segítségnyújtás, ellenőrzés, jóváhagyás stb.).	A béta teszt idején vizsgálni kell, hogy milyen erőforrásokat köt le a könyvtárosok oldalán a szolgáltatás működtetése, az adatok ellenőrzése.
Nem tudjuk kellő mértékben bevonni a belső munkatársakat az ellenőrzésbe, jóváhagyásba.	Megkeressük a közreműködők közül a legaktívabbakat, legtapasztaltabbakat és felkérjük őket, hogy az ellenőrzésben is legyenek segítségünkre.

A szövegjavítás, a kereshetőség és a szolgáltatás további kérdései:

- Hogyan mentjük el a változtatásokat? Hány változatot mentünk el? Melyiket tegyük elérhetővé?
- Hogyan tudjuk felmérni, hányan dolgoznak, mit csinálnak? Milyen statisztikát tudunk ehhez készíteni?
- Mérhető-e valamilyen módon a szöveg változása, a javítás eredményessége?
- Mit lásson javításkor a felhasználó: a nyers szöveget, a legutóbbi verziót, vagy mindent együtt?
- Ha sérülést, rongálást tapasztalunk, akkor melyik verzióhoz térünk vissza?
- A felhasználók végtelen számú alkalommal javíthatják a szöveget, vagy elérhet a szöveg egy végső, lezárt verziót?
- Ha egy gyengébb minőségű szöveg letárolódik, akkor minden korábbi javítás elvész?

Ezek nem könnyen megválaszolható kérdések. Tökéletesen biztonságos megoldást nem lehet kialakítani, de igyekszünk minimálisra csökkenteni a veszélyeket.

Hol tartunk?

A napokban elkészül egy olyan felület, ahol a kép kicsinyítve, de még olvasható szöveggel látszik. A képernyőn „fülek” segítségével választható az előzékek, a családnévek és a keresztnévek ábécérendes listája, ezekből kattintással lehet átemelni a keresett nevet az adatbázismezőkbe, és elemenként lehet „összerakni” belőle akár összetettebb személyneveket is. Az elhunyt nevét olyan módon rakjuk össze, ahogy az eredeti dokumentumon olvasható. A névelemekből kitüntetett szerepet kap a családnév és a keresztnév (asszonyok esetében a leánykori nevük), mert ennek alapján készül majd a böngészőlista, de a dokumentumhoz a teljes nevet hozzárendeljük.

A névmezőket szerkeszthetően jelenítjük meg, hogy javítható legyen a név, illetve új nevet is fel lehessen venni az adatbázisba, ha még nem fordult elő az elhunyt családnéve. Ebben az előkészítő fázisban a nevek kijelölésével és az adatbázisba helyezésével az adott gyászjelentés pontosan azonosítható lesz.

A második fázisban már összetettebb felületre van szükség, amelyen a gyászjelentés képe és nyers szövege együtt látszik, és a szöveg szerkeszthető, javítható formában jelenik meg. A kép is nagyítha-

tó lesz, s a nagyobb verziót külön ablakban lehet tanulmányozni. Cél a szövegek javítása, és egyúttal a releváns adatok átemelése a képernyőn megjelenített adatbázismezőkbe. A leíró adatok körét a könyvtáros szakemberek határozzák meg, megjelölve azt a néhány adatot, ami az eredményes, pontos kereséshez alapvetően szükséges.

Az előkészítés során az ábécé elejétől indulva sorban haladunk a képek azonosításával. Úgy gondoltuk, hogy amint így elkészül néhány ezer rekord, azokat áttesszük a második fázisba, vagyis ezekkel már tesztelhető lesz a javító felület működése. Elképzelhető, hogy a közreműködők között felosztjuk az anyagot, például filmenként haladunk, egy-egy filmnyi mennyiséget „kiadva” valakinek feldolgozásra. Azt is lehetővé szeretnénk tenni, hogy egy keresés eredményeként létrejött találati halmazt vesz „kézbe” valaki, és ezen halad a javítással. Utóbbi megoldás talán célravezetőbb, mert jobban közelíthetünk vele a segítők érdeklődési köréhez. Lehet, hogy valaki azért szeretne részt venni a munkában, mert saját családjára vonatkozó információkat keres. Így motiváltabb lesz, és javításaival ebben az esetben is segíti az adatbázis építését, az adatok javítását. Értelemszerűen a végső cél, hogy a teljes szöveg javított és kereshető állapotban legyen, valamint a fontos kutatási adatok egy adatbázisban, rendezetten, kereshetően tárolódjanak.

A digitális állomány bővítése

Természetesen szeretnénk azt is megoldani, hogy az interneten elérhetővé tett gyűjtemény tartalmilag is gazdagodjon. A jelenleg rendelkezésünkre álló digitális képek csak a 2003-ig beérkezett dokumentumokat tartalmazzák, de a kötelezpéldány-szolgáltatás keretében már jelentősen gyarapodott a nemzeti könyvtár gyűjteménye. Az utóbbi években beérkezett dokumentumokat is szeretnénk digitalizálni és hozzáférhetővé tenni. Terveink között szerepel, hogy a készülő szolgáltatás keretében lehetőséget nyújtsunk az önkéntes beadásra is, vagyis a könyvtár gyűjteményében nem szereplő (sok esetben nem is nyomdai úton sokszorosított) dokumentumok feltöltésére.

Tekintve a gyászjelentések nagy számát, komoly munka vár ránk, amelynek az időszükségletét még nem tudjuk megbecsülni. Az előrehaladás egyik fontos tényezője lesz, hogy hányan vállalkoznak arra, hogy részt vegyenek a munkában, és ki mennyi időt tud erre szánni. Az sem mellékes,

hogy a közreműködők mennyire pontosan, alaposan dolgoznak. Megkerülhetetlennek látszik, hogy az adatokat valaki ellenőrizze és jóváhagyja, erre is kell energiát, időt szánni. Célszerű lenne a felületen követni, hogy az adott rekord milyen javítottsági szintet ért el, illetve hogy jóváhagyott, kész rekordról van-e szó. A jóváhagyást könyvtárosok vagy erre a feladatra kijelölt segítők végeznék megfelelő ellenőrzés után.

Kérdés, hogy melyik fázisban érdemes szabadon, mindenki számára elérhetővé tenni a gyűjteményt, és a felhasználóknak mit mutassunk meg a javítás alatt lévő adatokból, szövegből.

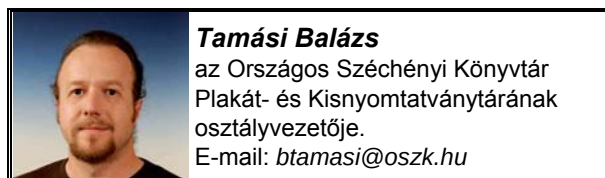
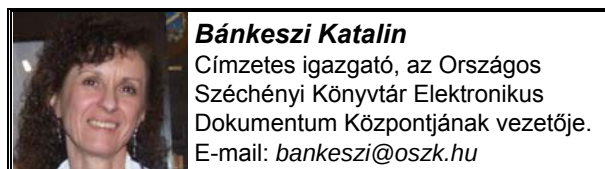
Célszerű lenne megvárni, amíg az előkészítő fázisban az elhunyt(ak) neve minden dokumentum esetében rendelkezésre áll, vagyis egyértelműen azonosíthatók a digitális képek. Amíg ez nincs így, addig a keresés csak megközelítő eredményt hoz, és lapozással lehet eljutni a keresett dokumentumhoz. Ha megvárjuk, amíg minden dokumentum egyedileg is megtalálható lesz, akkor néhány hónapos késéssel tudjuk a nagyközönség elé tárni a gyűjteményt. A javító fázis elkezdődhet előbb is, bár akkor még nem tud kiteljesedni, hiszen nem az egész állomány lesz elérhető, hanem csak a már azonosított részek. A javítással ekkor is jól lehet haladni, csak nem biztos, hogy megfelelő lesz a motiváltság, hiszen még messze nem minden név lesz kereshető.

A későbbiek során három jogosultsági szint kialakítását tervezzük. A legmélyebb hozzáférést az „admin” minősítésű közreműködők kapnak, akik az adatok ellenőrzésével és jóváhagyásával foglal-

koznak és öröködnék az adatbázis minőségén. A második szint a segítő közreműködők szintje, akik módosíthatnak az adatokon, javíthatják a szövegeket. A harmadik szint a felhasználóké, akik szabadon kereshetnek a már feltárt adatok között, illetve a teljes szövegben, de nincs lehetőségük módosítások végrehajtására. Nekik is lesz viszont lehetőségük arra, hogy elektronikus levélben, vagy magán a felületen jelezzék, ha hibát észlelnek.

A jelen cikk a tervezés előrehaladott szakaszában íródott, a fentiekben jelzett kérdések megválaszolása folyamatban van, ennek megfelelően alakul majd ki a szolgáltatás végleges koncepciója. Terveink szerint az elkövetkező hetekben megkezdődik a nevek pontos azonosítása, és még ebben az évben megjelenhet a javítófelület próbaverziója az első néhány ezer rekorddal. Reméljük, egy év múlva már beszámolhatunk az újszerű szolgáltatás bevezetésével kapcsolatos tapasztalatainkról és az első eredményekről.

Beérkezett: 2009. IX. 13-án.



A GPL-t és a Creative Commons licencet támogatja az Európai Unió

Az Európai Bizottság úgynevezett Zöld Könyvéhez kiadott állásfoglalásában az Európai Unió Gazdasági és Szociális Bizottsága egyértelműen a GPL és a Creative Commons licenc hatálya alá eső művek támogatásáról döntött.

A testület szerint a változtatások mielőbbi bevezetéséhez sürgősen módosítani kell a jelenleg hatályos szerzői jogi szabályokat és az EU szerzői jogi irányelvét. A módosítások célja, hogy megkönnyítsék a különböző könyvek és más tartalmak online kikölcsönzésének lehetőségét.

Az Európai Unió tavaly hozta nyilvánosságra Zöld Könyvét, amelynek célja a szerzői jog megreformálása, hogy az megfeleljen a 21. századi kutatás, a tudomány és az oktatás követelményeinek is. A döntéshozók azt szeretnék, hogy a tudás szabadon áramolhasson és ezáltal az eddignél több fejlesztésre kerülhessen sor az Európai Unió belső piacán. Szintén fontos egy, az EU mind a 27 tagállamában érvényes, egységes szerzői jogi környezet megteremtése. Biztosítanák azt is, hogy ha egy felhasználó nem férhet hozzá az információhoz, akkor fellebbezhessen a döntés ellen. Kiemelten támogatnák a GPL- és a Creative Commons licencű művek terjesztését, a két licenc elterjedését, valamint a GPL-licenc hatálya alá eső szoftverek alkalmazását.

Szintén kitér a dokumentum az úgynevezett elárult könyvekre, amelyeknél már nincsenek meg a szerzői jogi tulajdonosok, továbbá a felhasználók által készített tartalmakra és azok hasznosítására, illetve arra, hogy például a látássérültek vagy a vakok is az eddignél könnyebben férhessenek hozzá a különböző internetes anyagokhoz.

/SG.hu Hírlevél, 2009. október 5., <http://www.sg.hu/>

(SzP)